DOCUMENT RESUME

ED 252 907	CS 504 817
AUTHOR	Studdert-Kennedy, Michael, Ed.; O'Brien, Nancy, Ed.
TITLE	Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, July 1-December 31, 1984.
INSTITUTION	Haskins Labs., New Haven, Conn
SPONS AGENCY	National Institutes of Health (DHHS), Bethesda, Md.; National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Neurological and Communicative Disorders and Stroke (NIH), Bethesda, Md.; National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C.
REPORT NO	SR-79/80-(1984)
PUB DATE	84
CONTRACT	NICHHD-NO1-HD-1-2420; ONR-N00014-83-K-0083
GRANT	NICHHD-H-16591; NICHHD-HD-01994; NIHBRS-RR-05596; NINCDS-NS-13617; NINCDS-NS-13870; NINCDS-NS-18010; NSF-BNS-8111470
NOTE	258p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	Adults; *Articulation (Speech); *Communication Research; Elementary Secondary Education; Japanese; *Language Research; Learning Disabilities; *Listening Comprehension; *Phonetics; Reading Difficulties; Second Languages; Serbocroatian; *Speech Communication; *Speech Handicaps; Speech Pathology; Speech Skills; Spelling; Vowels

ABSTRACT

One in a series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical research applications, this document contains 17 research reports. Topics discussed in the reports are as follows: (1) dynamic modeling of phonetic structure; (2) coarticulation as a component in articulatory description; (3) contextual effects on lingual-mandibular coordination; (4) the timing of articulatory gestures; (5) the onset of voicing in stuttered and fluent utterances; (6) phonetic information integrated across intervening nonlinguistic sounds; (7) parameters of spectral/temporal fusion in speech perception; (8) monitoring for vowels in isolation and in a consonantal context; (9) the perception of /l/ and /r/ by native speakers of Japanese; (10) a theoretical note on speech timing; (11) reconciling monophthongal vowel percepts and continuously varying F patterns; (12) synergies--stabilities, instabilities, and modes; (13) repetition and comprehension of spoken sentences by reading disabled children; (14) spelling proficiency and sensitivity to word structure; (15) effects of phonological ambiguity on beginning readers of Serbo-Croatian; (16) verticality unparalleled; and (17) reiterant speech production. (HOD)

ED252907

SR-79/80 (1984)

Status Report on
SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 July - 31 December 1984

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06511

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

(The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Michael Studdert-Kennedy, Editor-in-Chief
Nancy O'Brien, Editor
Margo Carter, Technical Illustrator
Gail Reynolds, Technical Coordinator

ACKNOWLEDGMENTS

The research reported here was made possible
in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994
Grant HD-16591

National Institute of Child Health and Human Development
Contract NO1-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant BNS-8111470

National Institute of Neurological and Communicative
Disorders and Stroke
Grant NS 13870
Grant NS 13617
Grant NS 18010

Office of Naval Research
Contract N00014-83-K-0083

SR-79/80 (1984)
(July-December)

HASKINS LABORATORIES PERSONNEL IN SPEECH RESEARCH

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Thomas Baer
Patrice S. Beddor†
Fredericka Bell-Berti
Pier Marco Bertinetto[1]
Catherine Best*
Gloria J. Borden*
Susan Brady*
Catherine P. Browman
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Laurie B. Feldman*
Anne Fowler†
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Sarah Hawkins††
Daniel Holender[2]
Satoshi Horiguchi[3]
Leonard Katz*
J. A. Scott Kelso
Gary Kidd†
Andrea G. Levitt
Alvin M. Liberman*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Kevin Munhall†
Patrick W. Nye
Robert J. Porter[4]
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Mary Smith*
Michael Studdert-Kennedy
Betty Tuller*
Michael T. Turvey*
Ben C. Watson††
Douglas H. Whalen

Technical/Support

Michael Anstett
Margo Carter
Philip Chagnon
Alice Dadourian
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Sabina D. Koroluk
Bruce Martin
Betty J. Myers
Nancy O'Brien
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Dragana Barac
Sara Basson
Eric Bateson
Suzanne Boyce
Jo Ann Carlisle
Andre Cooper
Jan Edwards
Jo Estill
Nancy Fishbein
Carole E. Gelfer
Charles Hoequist
Bruce Kay
Noriko Kobayashi
Rena A. Krakow
Katrina Lukatela
Harriet Magen
Sharon Manuel
Richard McGowan
Susan Nittrouer
Lawrence D. Rosenblum
Hyla Rubin
Richard C. Schmidt
John Scholz
Robin Seider
Suzanne Smith
Katyanee Svastikula
Daniel Weiss
Deborah Wilkenfeld
David Williams

---

* Part-time
[1] Scuola Normale Superiore, Pisa, Italy
[2] Free University of Brussels, Brussels, Belgium
[3] Visiting from University of Tokyo, Japan
[4] Visiting from University of New Orleans and Kresge Research Laboratory of the South, New Orleans, Louisiana
† NIH Research Fellow
†† NRSA Training Fellow

CONTENTS

DYNAMIC MODELING OF PHONETIC STRUCTURE
Catherine P. Browman and Louis M. Goldstein 1-17

COARTICULATION AS A COMPONENT IN ARTICULATORY DESCRIPTION
Katherine S. Harris 19-37

CONTEXTUAL EFFECTS ON LINGUAL-MANDIBULAR COORDINATION
Jan Edwards 39-48

THE TIMING OF ARTICULATORY GESTURES: EVIDENCE FOR RELATIONAL INVARIANTS
Betty Tuller and J. A. Scott Kelso 49-64

ONSET OF VOICING IN STUTTERED AND FLUENT UTTERANCES
Gloria J. Borden, Thomas Baer, and Mary Kay Kenney 65-80

PHONETIC INFORMATION IS INTEGRATED ACROSS INTERVENING NONLINGUISTIC SOUNDS
D. H. Whalen and Arthur G. Samuel 81-92

PARAMETERS OF SPECTRAL/TEMPORAL FUSION IN SPEECH PERCEPTION
Bruno H. Repp and Shlomo Bentin 93-106

MONITORING FOR VOWELS IN ISOLATION AND IN A CONSONANTAL CONTEXT
Brad Rakerd, Robert R. Verbrugge, and
Donald P. Shankweiler 107-116

PERCEPTION OF [l] AND [r] BY NATIVE SPEAKERS OF JAPANESE: A DISTINCTION BETWEEN ARTICULATORY AND PHONETIC PERCEPTION
Virginia A. Mann 117-124

A QUALITATIVE DYNAMIC ANALYSIS OF REITERANT SPEECH PRODUCTION: PHASE PORTRAITS, KINEMATICS, AND DYNAMIC MODELING
J. A. S. Kelso, Eric Vatikiotis-Bateson,
Elliot L. Saltzman, and Bruce Kay 125-159

A THEORETICAL NOTE ON SPEECH TIMING
J. A. S. Kelso, Betty Tuller, and Katherine S. Harris 161-166

ON RECONCILING MONOPHTHONGAL VOWEL PERCEPTS AND CONTINUOUSLY VARYING F PATTERNS
Leigh Lisker 167-174

SYNERGIES: STABILITIES, INSTABILITIES, AND MODES
E. Saltzman and J. A. S. Kelso 175-179

REPETITION AND COMPREHENSION OF SPOKEN SENTENCES BY READING-DISABLED CHILDREN
Donald Shankweiler, Suzanne T. Smith, and Virginia A. Mann 181-196

SPELLING PROFICIENCY AND SENSITIVITY TO WORD STRUCTURE
F. William Fischer, Donald Shankweiler, and Isabelle Y. Liberman 197-221

EFFECTS OF PHONOLOGICAL AMBIGUITY ON BEGINNING READERS OF SERBO-CROATIAN
Laurie B. Feldman, G. Lukatela, and M. T. Turvey 223-240

VERTICALITY UNPARALLELED
Ignatius G. Mattingly and Alvin M. Liberman 241-245

PUBLICATIONS 249-250

APPENDIX: DTIC and ERIC numbers
(SR-21/22 - SR-79/80) 251-252

Status Report on Speech Research

Haskins Laboratories

DYNAMIC MODELING OF PHONETIC STRUCTURE*

Catherine P. Browman and Louis M. Goldstein†

Abstract. A dynamical approach to phonetics allows utterances to be represented as compact, linguistically-relevant structures that have inherent temporal, as well as spatial, properties. This conception of phonetic structure is exemplified and tested, in a preliminary way. Vertical movement of the lower lip in nonsense items of the form [bVbəbVb] was recorded. The vowels were either [i] or [a], and stress either initial or final. The measured articulatory trajectories were compared to the sinusoidal curves that would be generated by an undamped mass-spring dynamical system. The parameter values of the sinusoids were changed every quarter- or half-cycle, and were determined from the measured trajectories' durations and displacements. The frequency parameter was modulated every half-cycle, according to one of two alternate organizational hypotheses, consonant/vowel vs. transition. Both organizations modeled the trajectories closely, with a slight superiority for the consonant/vowel organization. Stress level had a systematic effect on the frequency parameter values, with stressed < unstressed < reduced (i.e., stressed lowest frequency). Word-initial stressed consonants were matched least well by the generated curves, suggesting the need for alternative dynamical models.

Much linguistic phonetic research has attempted to characterize phonetic units in terms of measurable physical parameters or features (Fant, 1973; Halle & Stevens, 1979; Jakobson, Fant, & Halle, 1951; Ladefoged, 1971). Basic to these approaches is the view that a phonetic description consists of a linear sequence of static physical measures--either articulatory configurations or acoustic parameters. The course of movement from one such configuration to another has been viewed as secondary. Recently, we have proposed (Browman & Goldstein, 1984) an alternative approach, one that characterizes phonetic structure as patterns of articulatory movement, or gestures, rather than static configurations. While the traditional approaches have viewed the continuous movement of vocal-tract articulators over time as "noise" that tends to obscure the segment-like structure of speech, we have argued that setting out to characterize articulator movement directly leads not to "noise," but to organized spatio-temporal structures that can be used as the

---

*In V. Fromkin (Ed.), Phonetic linguistics. New York: Academic Press, in press.

†Also Departments of Linguistics and Psychology, Yale University.

Acknowledgment. This research was supported by Grant No. NS-13617 from the National Institute of Neurological and Communicative Disorders and Stroke, and Grant No. HD-01994 from the National Institute of Child Health and Human Development.

basis for phonological generalizations as well as accurate physical description. In our view, then, a phonetic representation is a characterization of how a physical system (e.g., a vocal tract) changes over time. In this paper, we begin to explore the form that such a characterization could take, by attempting to explicitly model some observed articulatory trajectories.

Although we want to account for how articulators move over time, this does not mean that time per se must appear as a dimension of the description. In fact, a dimension of time would be quite problematic, because of temporal variations introduced by changes in speaking rate and stress. For example, suppose our phonetic description were to specify the positions of articulators at successive points in time. As speaking rate changes, the values at successive time points are all likely to change in rather complex ways. Such a representation would not, therefore, be very satisfactory. It would be preferable to describe phonetic structure as a system which produces behavior that is organized in time, but which does not require time as a control parameter (as has been suggested, for example, by Fowler, 1977, 1980). Like conventional phonetic representations, such a system does not explicitly refer to time. Unlike these representations, however, it explicitly generates patterns of articulator movement in time and space.

The dynamical approach to action currently being developed, e.g., by Kelso and Tuller (1984), and Saltzman and Kelso (1983), provides the kind of time-free structure that can characterize articulatory movement. The approach has been applied to certain aspects of speech production (Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Tuller, & Harris, 1983; Kelso, Tuller, & Harris, in press), as well as to more general aspects of motor coordination in biological systems (e.g., Kelso, Holt, Rubin, & Kugler, 1981; Kugler, Kelso, & Turvey, 1980). Previous approaches to motor coordination (e.g., Hollerbach, 1982) have emphasized the importance of a time-varying trajectory "plan" for the muscles and joints to follow in the performance of a coordinated activity, and require an intelligent executive to ensure that the plan is followed. In the dynamical approach taken by these investigators, actions are characterized by underlying dynamical systems, which once set into place, can autonomously regulate the activities of sets of muscles and joints over time.

A physical example of a dynamical system is a mass-spring system, that is, a movable object (mass) connected by a spring to some rigid support. If the mass is pulled, and the spring stretched beyond its equilibrium length, the mass will begin to oscillate. In the absence of friction, the equation characterizing motion is seen in (1), and the trajectory of the object attached to the spring can be seen in Figure 1.

$$m\ddot{x} + k(x - x_0) = 0 \qquad (1)$$

where m = mass of the object
k = stiffness of the spring
x_0 = rest length of the spring
x = instantaneous displacement of the object
$\ddot{x}$ = instantaneous acceleration of the object

Notice that an invariant organization (that in (1)) gives rise to the time-varying trajectory in Figure 1. No point-by-point plan is required to

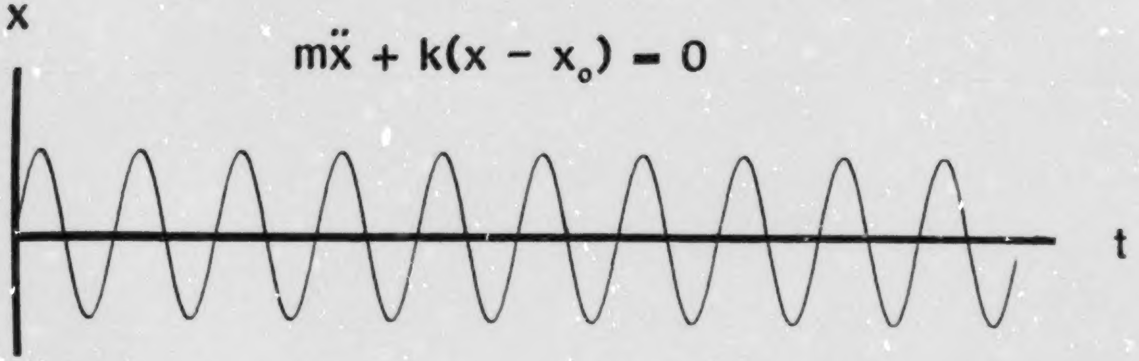


Figure 1. Output of undamped mass-spring system. Time (t) is on the abscissa and displacement (x) is on the ordinate.

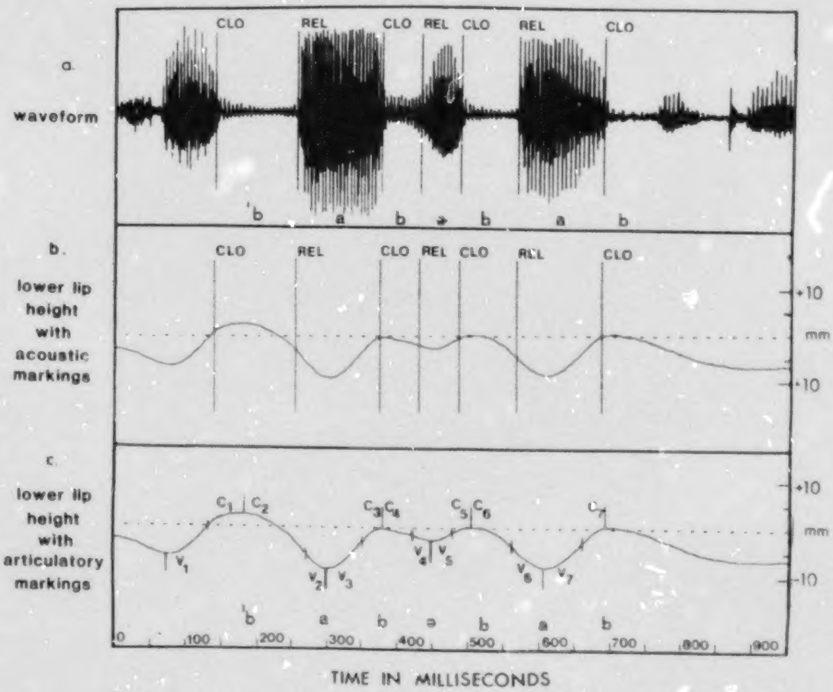


Figure 2. Lower lip height and waveform for single token of ['babəbab]. (a) Acoustic waveform with consonant closures and releases marked. (b) Lower lip height (in mm) over time. Closures and releases based on waveform are marked. (c) as in (b), but with tick marks to indicate displacement and velocity extrema. Intervals between extrema are labeled as discussed in text.

describe this pattern of movement, and time is not referred to explicitly. Only the parameter values and the initial conditions need be specified. The undamped mass-spring equation (1) is a very simple example of a dynamical system. It is important to note that this system can give rise to a whole family of trajectories, not just the one portrayed in Figure 1. Different trajectories can be generated by changing values of the system's parameters. For example, changing the stiffness of the spring will change the observed frequency of oscillation. Changes in the rest length of the spring and the initial displacement of the mass will affect the amplitude of the oscillations.

This simple mass-spring equation (generally with a linear or non-linear damping term added) exemplifies the dynamical approach to coordination and control of movement in biological systems in general, and of speech articulators in particular. The appeal of this approach lies both in its potentially simple description of articulatory movements (i.e., only a few underlying parameters serve to characterize a whole range of movements), and also in its physical and biological generality. In order to be useful for phonetics and linguistics, however, such a dynamical system must be related to phonetic structure. In one early attempt to specify this relationship, Lindblom (1967) proposed that a dynamical description could be used to account for speech duration data. More recently, Kelso, V.-Bateson, Saltzman, and Kay (in press) and Ostry, Keller, and Parush (1983) have analyzed variation in stress and speaking rate in terms of a dynamical model. In this paper, we explore a basic linguistic issue that arises in the attempt to couch phonetic representations in the language of dynamics, namely, the definitions of the articulatory gestures.

To begin to relate phonetic description to a dynamical system, let us consider a very simple example. Figure 2b shows the vertical position of a light-emitting diode (LED) on the lower lip of a speaker of American English, as she produces the utterance ['babəbab] in the frame "Say ____ again." The acoustic closures and releases marked on the articulatory trajectory are determined from the acoustic waveform, shown in Figure 2a. Note that the lower lip is raised (toward the upper lip) for the closures and lowered for the vowels. How can this observed lower-lip trajectory be described, in terms of a dynamical system? Clearly the lower lip is showing an oscillatory pattern, that is, it goes up and down in a fairly regular way, but it does not show the absolute regularity of our mass-spring system in Figure 1. For example, the lip is lower in the full vowels than in the schwa. Thus, a mass-spring organization with constant parameter values will not generate this lower lip trajectory. However, it might be possible to generate this kind of trajectory if the parameter values were changed in the course of the utterance. The underlying dynamical organization, together with the particular changes of the parameters, would then serve to characterize the phonetic structure of the utterance.

It is, of course, obvious that a characterization of lower lip position over time is not a complete phonetic representation. Nonetheless, in very simple utterances containing only bilabials and a single vowel, it comes quite close to being an adequate phonetic description. Browman, Goldstein, Kelso, Rubin, and Saltzman (1984) have shown that an alternating stress ['mama'mama...] sequence can be adequately synthesized using a vocal-tract simulation controlled by only two mass-spring systems--one for lip aperture (the distance between the two lips), and one for lip protrusion. Clearly, however, more complex utterances will require additional dynamical systems, and relationships among these systems; such interrelationships and their implications for phonology are discussed in Browman and Goldstein (1984). Even for the restricted utterances we will be considering here, we simplify the phonetic characterization by considering only the vertical position of the lower lip. We ignore horizontal lip displacement, the upper lip, and the fact that the movements of the lower lip can be decomposed into movements of the jaw and movements of the lower lip with respect to the jaw. The general framework we are operating within (the task dynamics of Saltzman & Kelso, 1983) allows us to describe the coordination of multi-articulator gestures, but this is irrelevant to the present paper, in which we consider only how to describe a particular articulator trajectory as the output of a dynamical system.

The undamped mass-spring system with constant parameter values generates sinusoidal trajectories with constant frequency and amplitude. We will show that observed trajectories can be directly modeled as sinusoids whose frequency and amplitude vary at particular points during the utterance. Of particular interest is how to define these points at which the values are changed. Since time is not a parameter of the system, they are defined not with respect to some reference clock, but rather in terms of the inherent cyclic properties of the dynamical system.

One set of inherently-definable points at which parameter values can be modulated are the points of minimum and maximum articulator displacement. Modulation at these points is suggested by studies of articulator movement that characterize trajectories in terms of opening and closing gestures (e.g., Kuehn & Moll, 1976; Parush, Ostry, & Munhall, 1983; Sussman, MacNeilage, & Hanson, 1973). Alternatively, points of peak velocity (both positive and negative) can also serve as dynamically-definable markers for modulation. In a simple mass-spring system, velocity peaks occur at the resting, or equilibrium, position. These different points of change imply different phonetic organizations, as can be seen with the help of Figure 2c. Here we see the same articulatory trajectory as in Figure 2b, with the addition of tick marks that indicate the displacement and velocity extrema. These points divide the utterance into intervals, each of which has been labeled either with a C (for consonant) or a V (for vowel). The consonant intervals are those on either side of a displacement peak, and the vowel intervals are those on either side of a displacement valley. Points of peak velocity, indicated by the smaller tick marks on the slopes, separate consonant intervals from vowel intervals. For example, V_1 is the interval from the minimum position of the lower lip in the frame vowel [ey], to the point of peak velocity as the lip starts to raise for the initial [b]. C_1 is the interval from this latter peak velocity point to the center of maximum lower lip height during the [b]. C_2 is the interval from this displacement peak to the peak velocity as the lower lip lowers for the following vowel [a].

If we change our model parameters only at displacement peaks and valleys, then successive VC or CV intervals (e.g., V_1-C_1, C_2-V_2) will be characterized with the same set of parameters. This constitutes a phonetic hypothesis that the articulatory trajectories can be modeled as successive CV and VC transition gestures, each with their characteristic values for the dynamical parameters. The parameters for these opening and closing gestures must take into account both the particular consonant and the particular vowel. Thus, this hypothesis provides a phonetic structure rather different from that commonly assumed in linguistics, in that it does not provide a physical characterization of individual consonants or vowels.

An alternative division of the articulatory trajectories is clearly possible, if we change parameters at velocity extrema rather than displacement extrema. In this way, successive C intervals will have a single characterization, as will successive V intervals (e.g., C_1-C_2, V_2-V_3). These new intervals, then, correspond roughly to consonant and vowel gestures, rather than to CV and VC transition gestures. Under this hypothesis, the relationship between the dynamical characterization and more conventional phonetic representations is somewhat more transparent than it is under the transition hypothesis. Note, however, that even under this hypothesis, consonants and vowels are defined in terms of dynamical structures, rather than as spatial targets.

In this paper, then, we present the results of some preliminary modeling of articulatory trajectories with sinusoids (the output of an undamped mass-spring system), under the C-V and transition hypotheses outlined above. In particular, the two hypotheses will be contrasted with respect to how the frequency parameter of the sinusoidal model is modulated. The frequency parameter (proportional to the square root of the stiffness of the underlying mass-spring system, assuming a unit mass) is of particular interest, because it controls the duration of a given gesture, and thus holds the key to how temporal (durational) regularities can be accommodated in a descriptive system that doesn't include time as a variable. Therefore, we will examine how the frequency of an articulatory gesture varies as a function of stress, position within the item, and vowel quality.

Method

Articulatory Trajectories

The trisyllabic nonsense items shown in Table 1 were chosen for analysis. Stress is either initial or final, with the second syllable always reduced, and the vowels are either [ɪ] or [a]. The items were recorded by a female speaker of American English in the carrier sentence "Say ____ again." Table 1 indicates the number of tokens of each of the items that were analyzed.

Table 1

Utterance	No. of Tokens
bibə'bib	11
'bibəbib	14
babə'bab	10
'babəbab	11

Movements of the talker's lips and jaw were tracked using a Selspot system that recorded displacements, in the mid-sagittal plane, of LEDs placed on the nose, upper lip, lower lip, and chin. The Selspot output was recorded on an FM tape recorder and was later digitally sampled at 200 Hz for computer analysis. To correct the articulator displacements for possible movements of the head, the Selspot signal for the nose LED was subtracted from each of the articulator signals. Each resulting articulator trajectory was then smoothed, using a 25 ms triangular window. For the present purpose, only the vertical displacement of the lower lip was analyzed.

Displacement maxima and minima were determined automatically using a peak-finding algorithm. Instantaneous velocities were computed by taking the difference of successive displacement samples. The maxima and minima of the resulting velocity curves were determined using the same program as for the displacements. Displacement and velocity extrema were used to divide each token into seven C and seven V intervals, as shown in Figure 2c.

Modeling

Each successive interval of each token was modeled as the output of a simple mass-spring system by fitting sinusoids to the articulatory trajectories. We generated the model trajectories using a sine-wave equation directly (equation (2)), in order to emphasize the inherent cyclic properties of dynamic systems. Recall that frequency is related to stiffness, and amplitude to rest length and maximum displacement. Thus, we controlled frequency, amplitude, and equilibrium position (rest length). (Phase is discussed below.) The individual model points--x'(j)--for an interval were generated according to equation (2), for the jth point in the interval (one point every 5 ms):

$$x'(j) = x_0 + A \sin(\omega k + \phi) \tag{2}$$

where x_0 = equilibrium position
A = amplitude
ω = frequency (in degrees per sample point)
ϕ = phase

Frequency varied every 2-interval gesture, where the gestures were defined according to the two hypotheses outlined in the previous section. For the C-V hypothesis, a gesture included the two intervals between successive velocity extrema (e.g., C_1-C_2, V_2-V_3). For the transition hypothesis, a gesture included the two intervals between successive displacement extrema (e.g., V_1-C_1, C_2-V_2). We posit that a gesture constitutes a half-cycle. Therefore, the frequency was computed as the reciprocal of twice the combined duration of the two intervals comprising a gesture. For example, the frequency used to model intervals C_1 and C_2 under the C-V hypothesis was 1 / (2 * (duration of C_1 + duration of C_2)). Similarly, the frequency for intervals C_2 and V_2 under the transition hypothesis was computed as 1 / (2 * (duration of C_2 + duration of V_2)).

Since our primary interest in this study was in the frequency parameter, we allowed the values of the equilibrium position and amplitude to change every interval. The values were determined from the initial and final displacement of the interval, adjusted for phase. The phase angle of a sine wave is 90 degrees at maximum displacement (the peaks), and 270 degrees at minimum displacement (the valleys). Amplitude and equilibrium position values were determined by the constraint that model and data agree exactly at these points, both in phase and in displacement. That is, the observed peaks and valleys were assumed to be the displacement extrema generated by the underlying model. The analogous assumption was not possible for the velocity extrema, however, since often the velocity extrema were not mid-way between the displacement extrema (as they would be if the parameter values weren't changing--cf. Figure 1). Thus, the observed velocity extrema did not correspond to 0 and 180 degrees in the modeled trajectories. Rather, the phases for these points in the model were permitted to vary according to the constraint that model and data agree exactly here as well as at the displacement extrema.

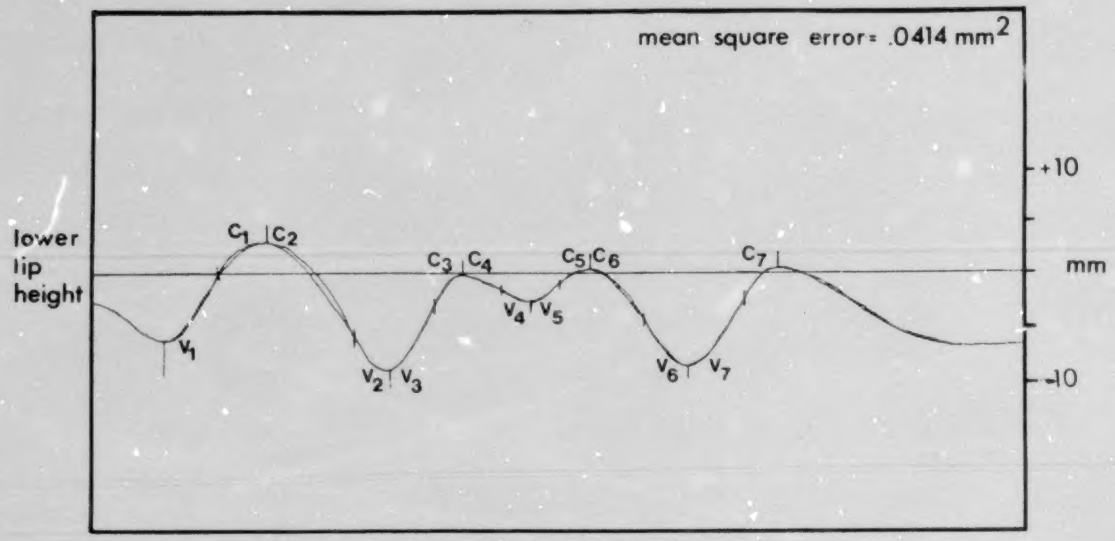
Results

Sinusoidal models are strikingly successful in fitting the articulatory data. Figure 3a shows the model trajectory generated for the C-V hypothesis superimposed on the real trajectory for our sample token of ['babəbab]. The curves lie almost completely on top of one another, diverging substantially only during the C_1, C_2, and C_6 intervals. This particular token is the best modeled of all ['babəbab] tokens, as measured by the mean square error of the modeled points. The token with the worst fit, not only for this utterance but for all the utterances, is shown in Figure 3b. Again, the curves lie almost completely on top of one another, diverging substantially only in the same places as in Figure 3a.

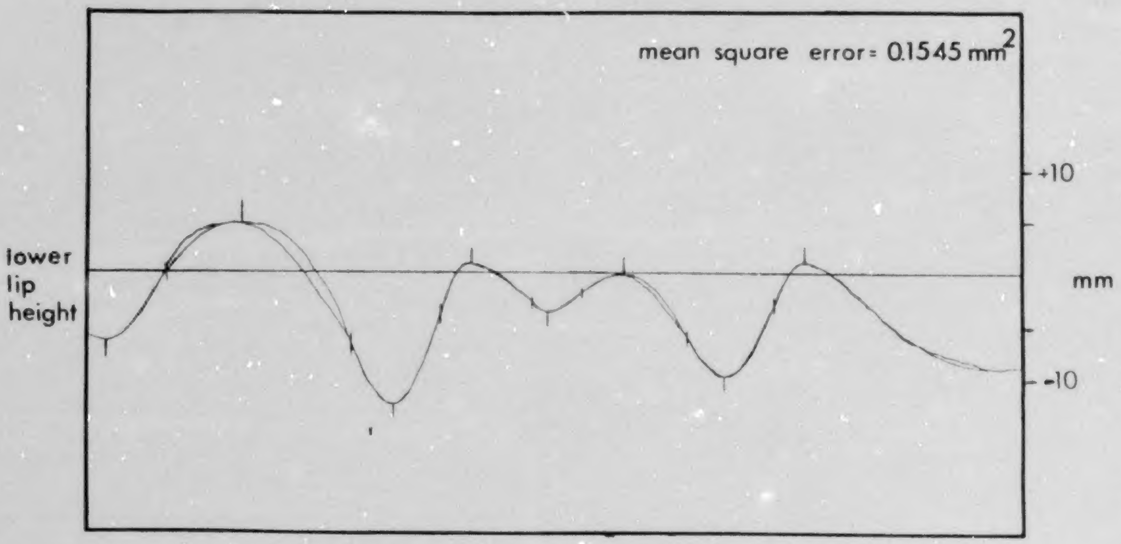
In general, the modeled trajectories for both hypotheses and for all utterances fit comparably to the trajectories shown in Figure 3a. Table 2 gives the mean square error averaged across all tokens for each of the four utterances under the C-V and transition hypotheses. The two hypotheses differ by only a small amount, but the C-V hypothesis appears to be consistently better. Comparison of individual tokens supports this slight superiority of dividing the trajectory into consonant and vowel gestures.

Table 2

Utterance	Mean Square Error (mm²)	
	C-V Hypothesis	Transition Hypothesis
bibə'bib	.0154	.0171
'bibəbib	.0466	.0615
babə'bab	.0358	.0471
'babəbab	.0907	.1220

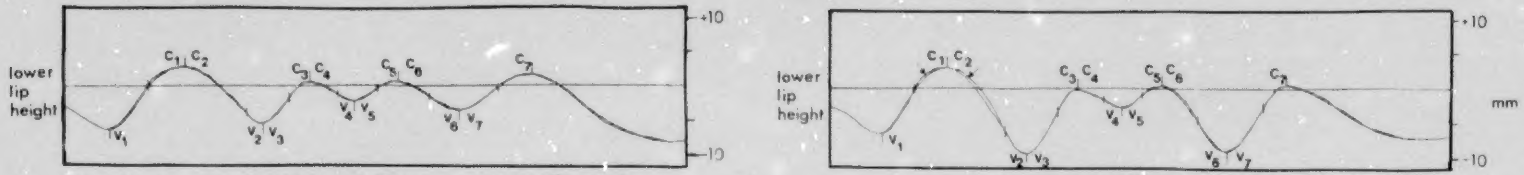


a. token with best fit for ['babəbab]



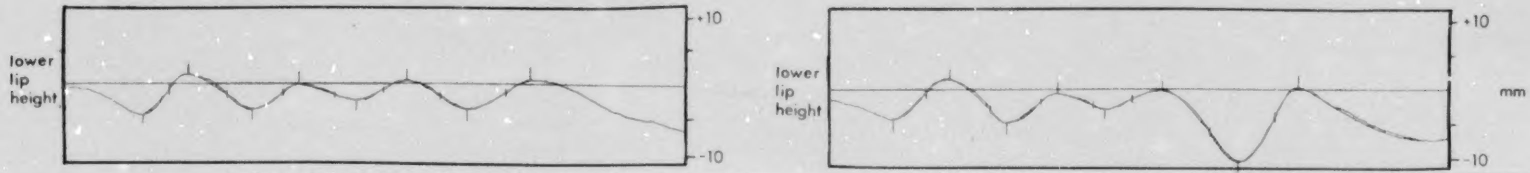
b. token with worst fit : ['babəbab]

Figure 3. Sample comparisons of superimposed model (C-V hypothesis) and data trajectories. (a) ['babəbab] token with the best fit. (b) token with the worst overall fit, which is also a token of ['babəbab].



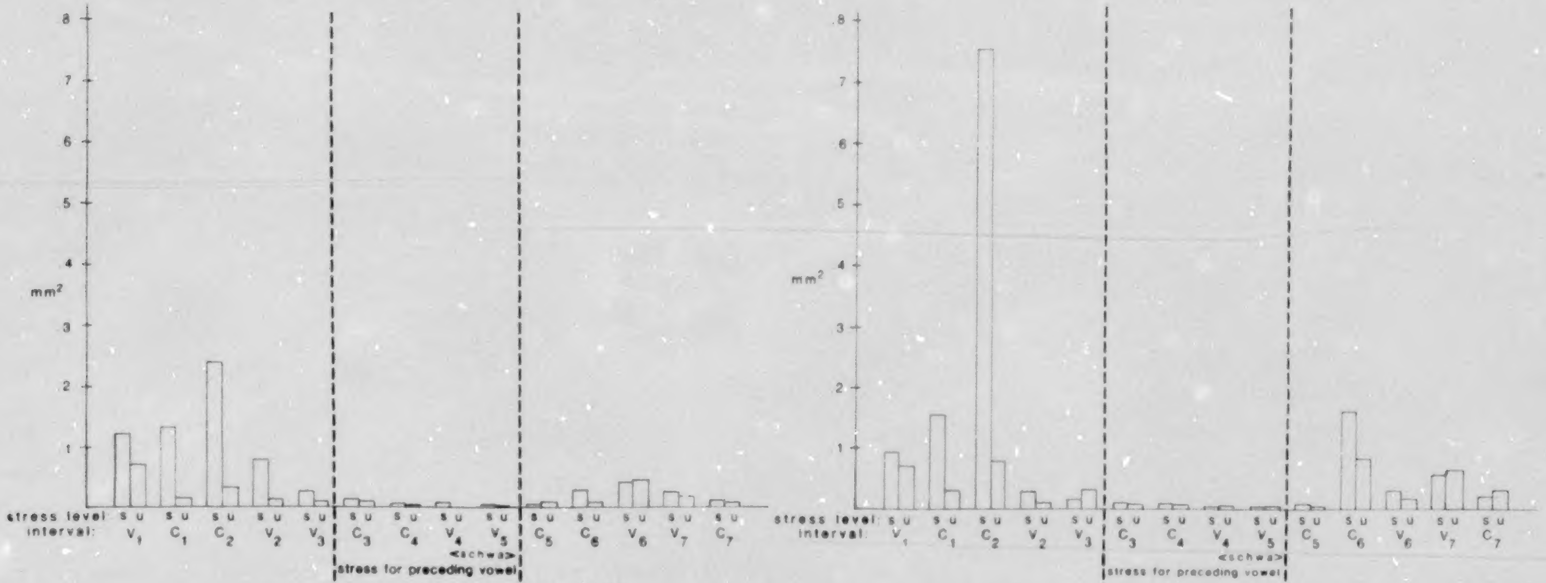
a. Superimposed data and model for ['bibəbib]

d. Superimposed data and model for ['babəbab]



b. Superimposed data and model for [bibə'bib]

e. Superimposed data and model for [babə'bab]



c. mean square error for items containing [i]

f. mean square error for items containing [a]

Figure 4. Best fit and error distributions by intervals for comparison of model (C-V hypothesis) and data trajectories. (a) Superimposed model and data for token of ['bibəbib] with best fit. (b) Superimposed model and data for token of [bibə'bib] with best fit. (c) Distribution of mean square error across trajectory intervals for items containing [i]. s=Stressed interval; u=Unstressed interval. For reduced intervals C3-V5, s=preceding vowel stressed, u=preceding vowel unstressed. (d) Superimposed model and data for token of ['babəbab] with best fit. Arrows indicate data. (e) Superimposed model and data for token of [babə'bab] with best fit. (f) Distribution of mean square error across trajectory intervals for items containing [a]. s=Stressed interval; u=Unstressed interval. For reduced intervals C3-V5, s=preceding vowel stressed, u=preceding vowel unstressed.

BEST COPY AVAILABLE

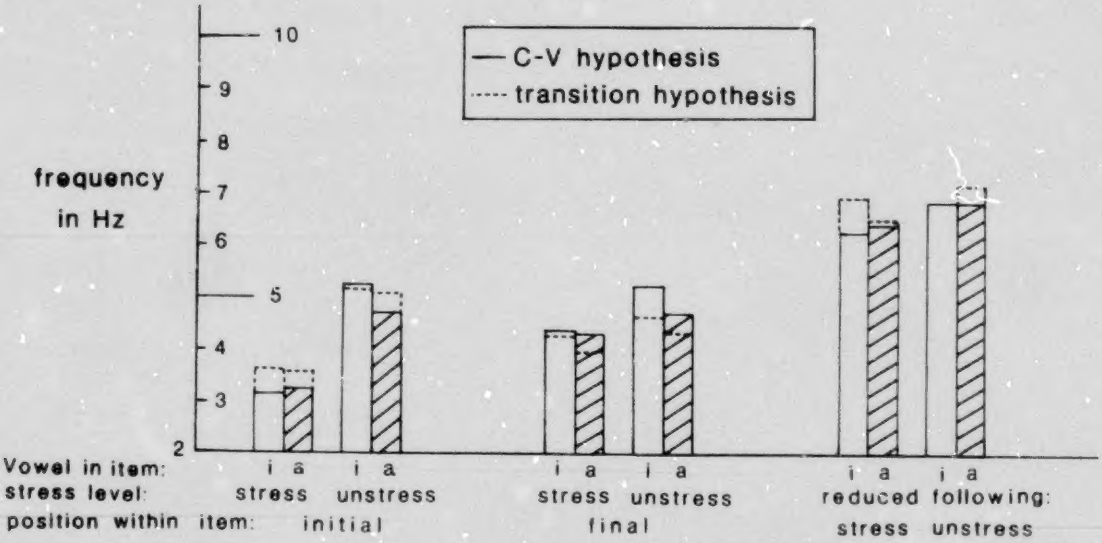
The contribution of different intervals of the trajectories to the error can be seen in Figure 4. The four curves show the model and data superimposed for the best tokens of each of the utterance types, under the C-V hypothesis. Utterances with [i] are shown on the left (a and b), and utterances with [a] are shown on the right (d and e). The graphs at the bottom of the figure show the mean square error for the individual intervals from V_1 to C_7. These are averages across all tokens of a given utterance. Again, results for utterances with [i] are shown at the left, and with [a] at the right. Intervals occurring in stressed and unstressed syllables are shown separately.

The error distributions show that the worst fit is found for item-initial stressed consonants, for both [a] and [i] utterances. In particular, interval C_2, the release of this initial stressed consonant, is poorly modeled relative to the other intervals. The release of the stressed consonant is also relatively poorly modeled in final syllables containing [a]. Examining the trajectories in the poorly modeled regions of ['babəbab] in Figure 4d, we can see that the actual consonant trajectory (indicated by arrows) shows a flatter top than that predicted by sinusoidal trajectories. This can, perhaps, be explained by noting that it tends to occur in regions in which the lower lip is raised quite high against the upper lip. The flattening may be the result of some limit on the compressibility of the lips. Alternatively, it may be that there is some tendency for initial stressed consonants to be "held," suggesting a somewhat different kind of dynamical system (e.g., a damped mass-spring).

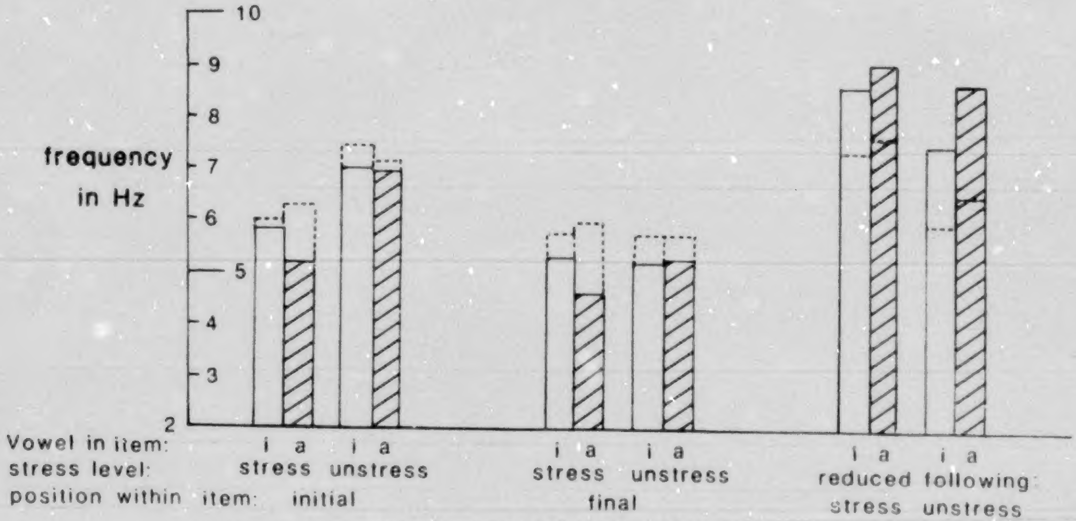
The error distributions also show a clear tendency for the reduced syllables to have the smallest error. This may partly be due to the fact that the actual displacement differences between the beginning and end of such intervals tend to be very small, and given that the ends are perfectly modeled, there simply isn't much room for error. Similarly, there is some tendency for utterances with [i] to show less error than utterances with [a]. Again, the lower lip shows less movement with [i] than [a], leaving less room for error. However, the smaller amplitude of movement does not completely account for the better fit. Correlations between amplitude of movement and error are not high, for example, .242 for [babə'bab]. Thus, the straight-forward mass-spring model we have chosen to investigate appears to be adequate for the unstressed and reduced syllables, but needs to be modified for stressed, item-initial consonants.

In addition to goodness-of-fit considerations, a dynamical phonetic structure can also be evaluated with respect to how well it can elucidate systematic variation. For example, we can examine how the values of the model parameters vary as a function of context. Given the preliminary nature of our modeling, we will simply show some easily observable trends, rather than present a detailed statistical analysis.

The bars with solid lines in Figure 5a show the mean value of the frequency parameter for the consonant gesture under the C-V hypothesis, as a function of the consonant's stress and position within the item for the two vowel contexts. Only the three consonants preceding vowels are shown. The first thing to note about the data is that the nature of the vowel in the item ([i] or [a]) has little effect on the consonant frequency (although unstressed [b] has a lower frequency before [a] than before [i]). That is, the consonant frequency is independent of vowel quality. Stress, however, clearly shows a systematic influence on the frequency of the [b] gesture. The consonant has a



a. consonant frequencies



b. vowel frequencies

Figure 5. Consonant and vowel frequencies generated by C-V and transition hypotheses, according to vowel in item, stress level, and position within item. For reduced consonants and vowels, always in medial position, stress or unstress refer to the preceding syllable.

higher frequency in unstressed syllables than in stressed syllables, for both the initial and final syllables in the item. In the medial reduced syllable, the consonant has the highest frequency of all. Kelso et al. (1984) also found unstressed gestures to be stiffer than stressed gestures, which is equivalent to an increase in frequency. This pattern of variation is completely consistent with the lengthening effect of stress as measured acoustically, (e.g., Klatt, 1976; Oller, 1973). Additionally, there is variation according to position. Item-initial stressed consonants are lower in frequency than consonants in the final syllable of the item. Again, this is consistent with observed acoustic word-initial consonant lengthening (Oller, 1973).

The vowel gestures are analyzed in a similar way in Figure 5b. Reduced vowels have higher frequencies than full vowels, as expected from the consonant data. Full vowels, however, do not behave quite as systematically as the consonants. For unstressed full vowels, there is little or no difference between [i] and [a] in frequency. Stressed full vowels, however, show a slight difference depending on whether the item contains [i] or [a]. Stressed [i] has a slightly higher frequency than stressed [a], which corresponds to the well-known acoustic duration difference noted, e.g., by Umeda (1975). (Reduced vowels show a possible compensatory effect, in that reduced vowels in items containing [i] have a lower frequency than those in items with [a].) The effect of stress for the full vowels is also not completely regular, but rather depends upon position. Only vowels in initial syllables show lower frequencies when stressed. Note, however, that vowels in final syllables are lower in frequency than those in initial syllables, which is in agreement with the acoustic effect of final lengthening (Klatt, 1975). It may be, then, that the final-lengthening effect washes out temporal differences between stressed and unstressed vowels in the final syllable. (At least one of Oller's (1973) subjects shows this kind of pattern.) Looked at in another way, when the initial vowel is stressed, it has about the same frequency as the unstressed final vowel in the same item. That is, the final lengthening effect is similar in magnitude to the stress effect. This is consistent with acoustic and perceptual investigations of stress patterns (Fry, 1958; Lea, 1977).

The bars with dotted lines superimposed on the solid-line bars in Figure 5 show the mean frequencies obtained under the transition hypothesis. For reasons to be discussed in the next section, the CV transitional gestures have been superimposed on the corresponding C gestures, and the VC transitions on the corresponding V gestures. (For example, the consonant in initial position, which represents the the consonant closing and release (C_1-C_2) under the C-V hypothesis, represents, under the transition hypothesis, the consonant release and movement to the following vowel (C_2-V_2). Similarly, the initial vowel represents V_2-V_3 under the C-V hypothesis and V_3-C_3 under the transition hypothesis.) Comparison of the dotted lines and solid bars shows substantial similarity. The only important differences are in the frequencies of the reduced vowels, which in the transition hypothesis are not higher than the full vowels. This is perhaps not surprising, given that the VCs that constitute the reduced syllables (V_5-C_5) include the initial consonant interval (C_5) of the following unreduced syllable.

To summarize, both the C-V hypothesis and the transition hypothesis fit the data quite well (except for stressed item-initial consonants), and generate very similar frequencies. The two hypotheses differ slightly in that the C-V hypothesis provides marginally better fit, and they predict differing patterns of frequencies for reduced vowels. Only stressed and reduced vowels show a difference in the frequencies generated for items containing [a] and items containing [i]. Stress level, however, has a generally consistent effect, with stressed syllables having the lowest frequency, unstressed syllables somewhat higher, and syllables containing reduced vowels having the highest frequency. This stress effect fails only for full vowels in final syllables, which in addition display lowered frequencies relative to initial syllables. Consonants, in contrast, have lower frequencies in initial syllables than in final. These stress and position effects are consistent with acoustic duration effects noted in the literature. Thus, well-known aspects of the temporal organization of speech can be accounted for in a model that does not explicitly refer to time.

Implications and Prospects

The success of a very simple dynamical system in modeling the observed trajectories of individual gestures gives important empirical support to the dynamical approach to phonetic structure. The approach is theoretically appealing because it provides a way of explicitly generating articulator trajectories from a time-free sequence of parameter specifications for consonants and vowels. This is made possible by recognizing, as suggested by Fowler (1977, 1980) and Fowler et al. (1980), that a phonetic structure is not just a linear sequence of parameter, or feature, values, but also must be described as some particular dynamical organization that the parameter values serve to modulate. The successive changes in parameter values can be linked to particular points in the underlying dynamical organization. This differs from conventional phonetic representations that do not provide any explicit way of generating articulatory trajectories from a sequence of parameter specifications.

The present model is only a preliminary validation of the general approach. A number of improvements need to be made before it can be claimed to have predictive power. In particular, the interval-by-interval specification of amplitude, with end-points exactly matched, needs to be replaced with a procedure that allows amplitude to be specified over longer stretches. The determination of frequency should be made in a way that is less vulnerable to experimental (and theoretical) error in determining the end-points of the gestures. Both frequency and amplitude should ultimately be determined by general linguistic parameters, for example, stress level and position, rather than by item-specific trajectory matching. These improvements can be carried out using the present simple undamped mass-spring dynamical model. In addition, alternative dynamical models need to be explored, in order to account for the poorly-matched item-initial stressed consonants, as well as inter-articulator compensation effects (cf. Saltzman & Kelso, 1983).

Another area to be investigated further is the organization of the underlying phonetic structure. This paper compared two organizational hypotheses, consonant-vowel gestures, and transitional gestures. While both hypotheses fit the data quite well, in this preliminary test, there is some indication that additional organizational hypotheses should be explored in future modeling attempts.

In the comparison of the two hypotheses, the CV transition was equated with the C, and the VC transition with the V. This was a post-hoc decision, based on the similarity of frequencies when the two hypotheses were so equated. In fact, the frequencies would not appear similar at all if the CV transitions were equated with Vs, rather than with Cs, and the VC transitions were likewise switched. Why the frequencies should line up this way is not clear. It may simply be the case that the intervals immediately following the displacement extrema (which are the intervals common to the C (or V) gestures and their equated transition gestures) are those in which frequency is crucially controlled. This interpretation is supported by results from an additional analysis, in which frequency was determined interval by interval, rather than by using two contiguous intervals. In this analysis, exactly those intervals following the displacement extrema displayed the stress and position patterns discussed in the preceding section, while the alternate intervals showed no clear relationship to the linguistic variables. However, there is also a more interesting account. This involves positing a structure in which frequency is fixed over a larger span of at least three intervals, e.g., C_1-C_2-V_2 and V_2-V_3-C_3. These longer gestures constitute a kind of overlapping organization (V_2 appears in both above), which is independently motivated by the kinds of coarticulatory phenomena typically observed in speech (cf. the overlapping segment analysis of coarticulation presented by Fowler, 1983).

Some such concept of overlapping gestures is also suggested by another regularity observable in the frequency patterns. The frequency of a consonant gesture under the C-V hypothesis is lower than the frequency of the vowel that follows it. This is counter to the common assumption that consonants involve short, rapid movements, while full vowels correspond to longer movements. The common assumption might, of course, be wrong. But such a counter-intuitive result may also be indicative of methodological problems, such as the choice of end-points, or of a basic flaw in the hypothesis generating the result. One obvious candidate for such a flaw is the assumption, in both hypotheses investigated, of independent, sequential gestures. Such an assumption was useful as a starting point, but is unlikely to be accurate. Rather, some form of overlap of the gestures--coarticulation--would likely give a better picture, and will be permitted in future modeling attempts. A possible overlapping structure is one in which consonantal gestures are phased relative to ongoing vowel gestures (cf. Tuller, Kelso, & Harris, 1982).

Finally, the comparison of the C-V hypothesis with the transitional hypothesis carries certain implications, not only for future research into phonetic organization, but also for the interpretation of past studies. Investigations into the nature of speech articulator movements have tacitly assumed the transition hypothesis (e.g., Kuehn & Moll, 1976; Parush et al., 1983; Sussman et al., 1973), and have consequently couched the description of their results in terms of opening and closing gestures. The present study, however, shows that the C-V hypothesis provides an organization that captures all of the same generalizations in the data as the transitional hypothesis; one that fits the data as well as or better than the transitional hypothesis; and moreover, one that is more immediately relatable to traditional linguistic units. In addition, while the two hypotheses generally produce equivalent frequency analyses, in at least one case--that of reduced vowels--they appear to differ substantively. The present study does not constitute evidence for one hypothesis over the other, given the overall similarity in fit. However, it does constitute evidence that the C-V organization, or some variant thereof, warrants serious consideration in the interpretation of speech articulator movement data. In general, we think that bringing dynamical principles to bear on problems of linguistic organization will lead to more linguistically-relevant accounts of speech production, as well as to a much richer, yet simple, conception of phonetic structure. The structure comprises an underlying dynamical system with associated parameter values. Together, the system and its parameters explicitly generate patterns of articulator movement. In addition, as we have demonstrated, such structures can retain the useful descriptive properties of more conventional phonetic representations.

References

Browman, C. P., & Goldstein, L. (1984). Towards an articulatory phonology. Unpublished manuscript.

Browman, C. P., Goldstein, L., Kelso, J. A. S., Rubin, P., & Saltzman, E. (1984). Articulatory synthesis from underlying dynamics. Journal of the Acoustical Society of America, 75, S22-23. (Abstract).

Fant, G. (1973). Distinctive features and phonetic dimensions. In G. Fant (Ed.), Speech sounds and features (pp. 171-191). Cambridge, MA: MIT Press. (Originally published 1969)

Fowler, C. A. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing control. Journal of Phonetics, 8, 113-133.

Fowler, C. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. Journal of Experimental Psychology: General, 112, 386-412.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. New York: Academic Press.

Fry, D. B. (1958). Experiments in the perception of stress. Language and Speech, 1, 126-152.

Halle, M., & Stevens, K. N. (1979). Some reflections on the theoretical bases of phonetics. In B. Lindblom & S. Öhman (Eds.), Frontiers of speech communication research (pp. 335-353). New York: Academic Press.

Hollerbach, J. M. (1982). Computers, brains, and the control of movement. Trends in Neuroscience, 5, 189-192.

Jakobson, R., Fant, C. G. M., & Halle, M. (1951). Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, MA: MIT.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.

Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of neuroscience (pp. 321-356). New York: Plenum.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (in press). A theoretical note on speech timing. In J. S. Perkell & D. Klatt (Eds.), Invariance and variation in speech processes. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E. L., & Kay, B. (in press). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. Journal of Phonetics, 3, 129-140.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59, 1208-1221.

Kuehn, D. R., & Moll, K. L. (1976). A cineradiographic study of VC and CV articulatory velocities. Journal of Phonetics, 4, 303-320.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Ladefoged, P. (1971). Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

Lea, W. A. (1977). Acoustic correlates of stress and juncture. In L. M. Hyman (Ed.), Studies in stress and accent. Los Angeles: University of Southern California.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. Speech Transmission Laboratory Quarterly Progress Report, STL-QPSR-4, 1-29.

Oller, D. (1973). The effects of position in utterance on speech segment duration. Journal of the Acoustical Society of America, 54, 1235-1246.

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 622-636.

Parush, A., Ostry, D. J., & Munhall, K. G. (1983). A kinematic study of lingual coarticulation in VCV sequences. Journal of the Acoustical Society of America, 74, 1115-1125.

Saltzman, E. L., & Kelso, J. A. S., (1983). Skilled actions: A task dynamic approach. Haskins Laboratories Status Report on Speech Research, SR-76, 3-50.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of labial consonants: Preliminary observations. Journal of Speech and Hearing Research, 16, 397-420.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 8, 460-472.

Umeda, N. (1975). Vowel duration in American English. Journal of the Acoustical Society of America, 58, 434-455.

COARTICULATION AS A COMPONENT IN ARTICULATORY DESCRIPTION*

Katherine S. Harris†

Coarticulation in Conventional Descriptions

In the recent past, the speech pathologist was often given a course in "articulatory phonetics." This study had as its goal teaching the student to make a series of alphabetlike symbols on a piece of paper, which, if the training was successful, would enable the student to perform such tricks as to read aloud in the "dialect" of the original speaker. Indeed, in the academic setting where this form was most highly developed--London University, the home base of Henry Sweet--these methods were used to change speech patterns not only of countless cockneys but also of the many non-native speakers of English who swarmed to London in the great days of the British empire. Of course, Sweet was the historical model for the hero of the play _Pygmalion_ and the musical _My Fair Lady_ (Borden & Harris, 1980).

In such training schemes, it was routinely assumed that there was no great difficulty about producing an adequate representation of the detailed act of speaking from alphabetlike marks on the page in which the only representation of time was the indication of visual succession (Lisker, 1974). Even now, it may be debated whether our knowledge of the principles of alphabetic writing is what underlies a belief in the adequacy of the symbol-by-symbol representation of speech, or whether, alternatively, the principles of alphabetic writing depend on some property of the perceptual system that makes such a representation seem adequate. Whichever formulation one prefers, there is a long history of a relationship between the study of phonetics and the desire of various authors, at various times, to commit oral narratives to writing. For example, as long ago as the 12th century, an Icelandic scholar wrote the "first grammatical treatise," an attempt to rework the orthography of Roman writing to suit the demands of representing the sounds of his native tongue (Fischer-Jørgenson, 1975).

The assumption that a series of symbols is an adequate representation of a child's articulation is one of the two basic assumptions of the typical course taken by the speech pathologist. The other is that, by listening, the transcriber can infer articulation or, at least, that aspect of articulation that is frequently all that the course provides--a schematic lateral view of the steady-state position of the articulators, to be associated with the left-to-right alphabetic labels of transcription.

---

*Also in R. G. Daniloff (Ed.), _Articulatory assessment and treatment issues_. San Diego, CA: College-Hill Press, 1984.

†Also Graduate School, City University of New York.

Acknowledgment. This work was supported by NINCDS Grant NS-13617 and NIH Biomedical Research Grant RR-05596 to Haskins Laboratories.

While the skills of a well-trained phonetician to reproduce speech are often astonishing to the layman, it is not clear what information is being used in performing. Even very well-trained phoneticians may not do a very good job of judging the articulator position associated with a given phone. For example, Ladefoged (1967) has shown that London-trained phoneticians cannot accurately assign tongue positions to the "cardinal vowels" produced by their London-trained colleagues. (The cardinal vowels are a reference system of articulator positions that give a kind of grid for vowels.) Indeed, it is his contention that vowels are sorted into categories on the basis of acoustic rather than articulatory similarity. In part, the phonetician's difficulty in making articulator position inferences is the inevitable result of the asymmetry of the relationship between acoustics and articulator position. Theoretically, although the acoustic signal can be estimated from a sufficiently detailed knowledge of vocal tract shape, a given acoustic signal may be associated with any of an infinite number of vocal tract shapes. An amusing example is provided by Ladefoged (Ladefoged, Harshman, Goldstein, & Rice, 1978). He shows two lateral views of the vocal tract. In one view, the vocal tract has a physiologically sensible contour. In the other, the tongue appears to have been creased into pleats. The two shapes are acoustically indistinguishable.

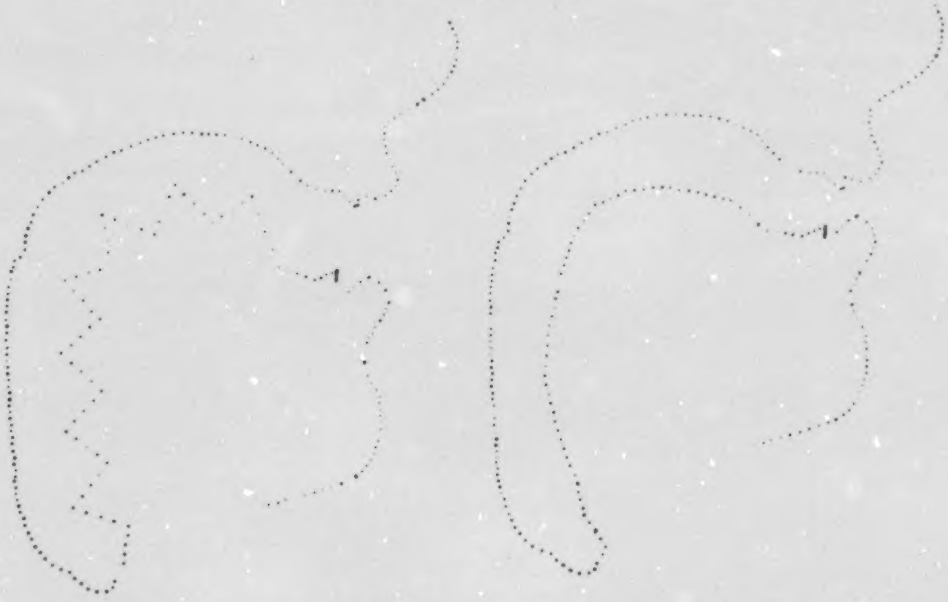


Figure 1. Two vocal tract shapes which generate the same formant values. Reproduced from Ladefoged, P., et al., 1978, op. cit.

In many of the more recent versions of our hypothetical course in articulatory phonetics, it is suggested that students learn to transcribe in "feature" notation, although a number of alternative descriptions that fit this rubric have been proposed (e.g., Chomsky & Halle, 1968; Ladefoged, 1971; Singh, 1978). It is not my interest here to attack the value of feature descriptions in principle. Within speech pathology as a field, they are useful in describing such diverse phenomena as the confusion matrices generated by hearing-impaired listeners in speech perception studies (Bilger & Wang, 1976) and the transfer of training in articulation correction (Compton, 1976; Pollack & Rees, 1972). The classic feature description is temporally isomorphic with a phonological description. The feature description, in its most sophisticated form (Chomsky & Halle, 1968), was developed to capture certain kinds of generalization within linguistics, such as morphophonemic alternation rules; the fact that the features have a physiological referent is not, in principle, an issue within the generative phonology framework. From the point of view of temporal structure, the features are abstract and timeless in the same sense as the units they were designed to replace.

The picture of speech production that our hypothetical student might infer, then, would be that the act of speaking proceeds from steady state to steady state, with (since the articulators must move continuously) some uninteresting events between, and that the articulatory origins of the steady state events are fairly transparent.

For many members of the research community, the sheer conspicuousness of the dynamic, as contrasted to the static, characteristics of the speech signal was first revealed by the illustrations in the book *Visible Speech* (Potter, Kopp, & Green, 1947), when it appeared shortly after the Second World War. The book represented, in many ways, the culmination of efforts by the Bell Telephone Laboratories to execute a mission inherited from Alexander Graham Bell himself. Bell had an interest both in the visual representation of speech and in using this representation to aid the deaf in learning to talk (Borden & Harris, 1980; Bruce, 1973). The attitude taken by Potter, Kopp, and Green towards the temporal structure is an interesting one, given their pedagogical purpose; one must learn to recognize the "characteristic position" or "hub," and the coarticulatory influences on it. While mention was (necessarily) made of the time-varying nature of the pattern, they said almost nothing about the time course of events as characteristics of speech sound representation. In other words, they took a segmental approach, although the dynamic aspects of the pattern were quite conspicuous.

It is a mistake to suppose that phoneticians whose main work preceded the sound spectrograph were wholly unaware of temporal phenomena, although these phenomena fit uneasily into any transcriptional description. For example, diphthongs are conventionally transcribed with two symbols, although their dynamic character was recognized. Jones, Sweet's successor, said: "For the purpose of practical language teaching it is convenient to regard a diphthong as a succession of two vowels, in spite of the fact that, strictly speaking, it is a gliding sound" (Jones, 1956, p. 99).

Earlier phoneticians were also well aware of the consequences to articulator movement; articulator position for one sound might influence that for a temporally adjacent one. This is the phenomenon called *assimilation* by Jones. For a common example, in the pronunciation of *these shoes in ordinary* speech, the /z/ /ʃ/ sequence is reduced to a single tongue movement to provide a suit-able position for /ʃ/. However, phoneticians were unaware of the extent to which the phenomena described above were not special examples; that is, since articulator position changes continuously, context sensitivity is the rule, rather than a phenomenon to be explained in special cases. Much of the effort for the following decade in the study of both speech production and speech perception was to build theories to account for the mismatch, in perception and production, between transcriptional phonetics and the phenomena of speech production. Theories in this field may be divided into two broad classes--we might call them discrete and continuous.

Theories of Speech Production

In this section, we will discuss some fairly recent speech production and perception theories. To a certain degree, these theories were aimed at rationalizing transcriptional or perceptual simplicity in the face of acoustic or articulatory variability. As noted above, the theories are of two basic kinds: discrete and continuous.

As an example of a discrete model, one might choose Perkell's model of the speech production process (Perkell, 1980), which is, in turn, based on Stevens' quantal model (1973). Without going into the details of the model shown in Figure 2, it can be seen to have stages such that the input, at the top of the figure, is a series of segments (S_1, S_2, S_3, and S_4) with each segment specified by a feature matrix, which is transformed into an isomorphic sensory goal. In the output, due to various hypothesized mechanisms, the boundaries between segments are no longer perpendicular lines, so that the "motor goals" and the segments are no longer isomorphic. This model is very like the one proposed by Henke (1966) to explain coarticulation, which will be discussed below. Two points should be noted: the only representation of time in the input is a simple succession, as in transcription, and the effect of reorganization is to desynchronize the representations of the transcriptional units.

An alternative point of view, although in very primitive form, is represented by Liberman's motor theory (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The motor theory is designed to account for the finding that two acoustic synthetic speech patterns will both produce the perceptual impression of the same consonant /d/, coupled with two different vowels (see Figure 3). Apparently the percept depends in some rather direct way on the dynamics of the acoustic pattern. The motor theory assumed that the listener must perform some operation dependent on the articulatory dynamics of production. This is a continuous theory because the dynamics of the pattern are important in themselves. It should be pointed out, however, that while the motor theory can be described as a continuous theory, Liberman has produced a stage model of the production-perception process that is quite similar to Perkell's (Liberman, 1970), and that Perkell, while in this classification a discrete theorist, has produced an extremely elegant discussion of articulatory dynamics from a quantal theory perspective (Perkell & Nelson, 1982).

It should be noted here, as well, that there is an apparent dichotomy between theories with some kind of linguistic referent, as discrete, and theories with some kind of motor referent, as continuous. This dichotomy is certainly not a necessary one. Thus, Fowler has argued (1977) that although symbols for speech may be represented in a particular form on a page, this does not mean that their motor representations take the same form in the ner-

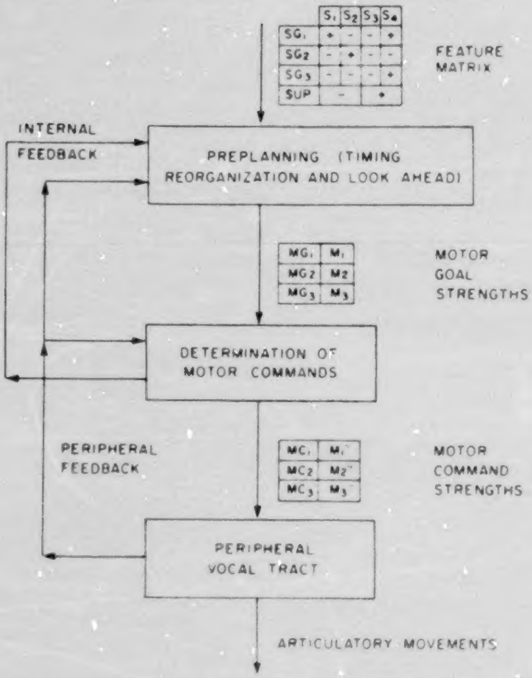


Figure 2. A figure showing Perkell's model of the translation of a feature matrix representation into articulatory units. Reproduced from Perkell, J., 1980, op. cit.

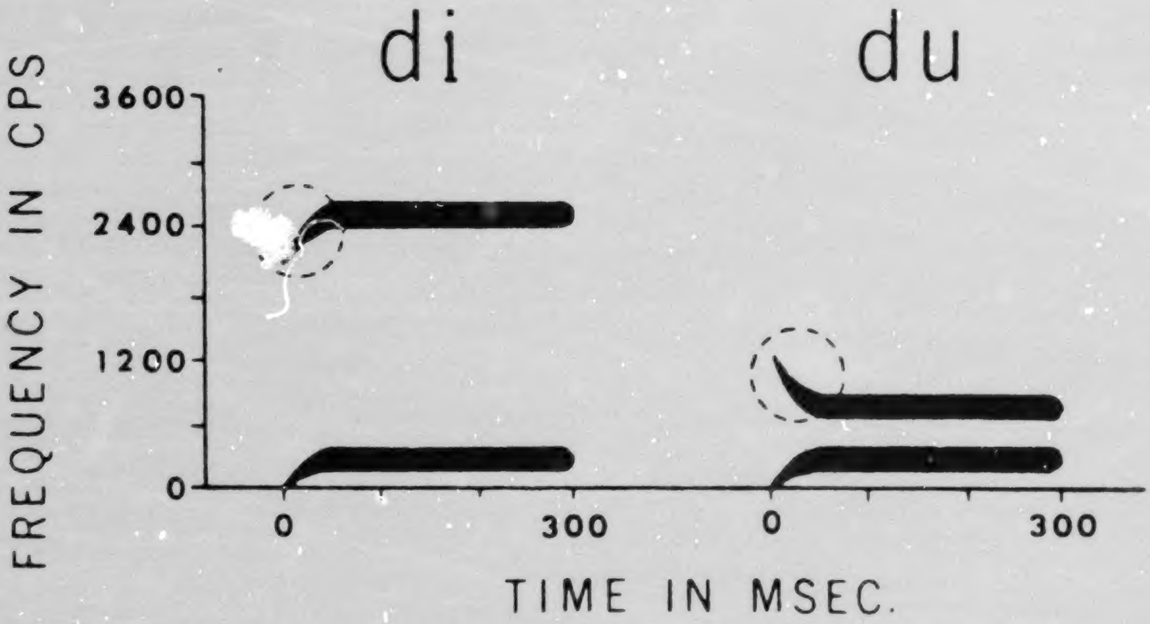


Figure 3. Two patterns perceived as /d/, followed by different vowels. Reproduced from Liberman, A., 1970, op. cit.

vous system. It is possible to argue that a representation of a motor plan in which the speech act is conceived as present somewhere in the nervous system, stripped of its temporal properties, which are then added in the execution, is very like the conception of speech as a phonological string.

It is important when examining theories of coarticulation in detail, as we shall do below, to recognize that the study of coarticulation is merely a small part of the study of skilled movement. Speech is special, as a type of skilled movement, in some rather unfortunate ways. For one thing, as we discussed above, speech comes with a notation scheme developed for special purposes, which may lead us astray when we attempt a more physiological description (Moll, Zimmermann, & Smith, 1977). Speech comes, as well, with a very inaccessible set of independent variables, as most articulators are difficult or impossible to observe without special techniques. However, even if experimental data on the movement of the articulators were easily gathered, one could not develop a theory of coarticulation simply by turning to a formulation lying ready-made in, for example, robotics. Machines can be produced that will mimic particular acts, but machines cannot now be designed that will adapt to a wide variety of changed environmental conditions, as humans do (Kelso, 1981). Furthermore, while we know a great deal about the muscular and neurological structures that participate in movement, the increase in our knowledge of structure does not help us very much with respect to function. For example, although a recent review chapter (Matthews, 1981) testifies to the explosion of our knowledge of the microstructure of the muscle spindle, a specialized device that provides feedback information about movement, the basic behavioral questions we ask about movement today are not very different from those we asked in the early 1930s, when Bernstein began his studies of the coordination of gait (summarized in Bernstein, 1967) or, perhaps, even when Sherrington summarized his observations of the decerebrate cat (Sherrington, 1906). We still lack a comprehensive theory that explains why skilled movements can be scaled up and down in timing, what causes the resistance of movement patterns to disruption by environmental change, and, with reference to coarticulation, why the elements of skilled movement patterns can be so freely reassembled to form novel sequences. While we have theories of coarticulation, as we will see below, they can rather easily be shown to fail. In what follows, I will attempt to outline the proposals for a model of coarticulation and to show how existing data succeed or fail in supporting them.

Hypthoses about Organizational Units and Speech Planning

Coarticulation as conventionally described is but one of a number of phenomena indicating some kind of organizational cohesiveness in speech. A great deal of effort has been directed at defining the outer bound over which such organizational cohesiveness exists. Unfortunately, the larger the unit that has been investigated, the larger the unit over which organizational dependencies can be demonstrated. For example, Lehiste has shown evidence for paragraph cohesivness over units that are larger than sentences. Speakers apparently signal first and last sentences in paragraphs by a number of means. The initial sentence in a paragraph is often signaled by high fundamental frequency, the last sentence by low fundamental frequency and laryngealization. There are, in addition, durational cues for the termination of paragraphs, although the way duration is used is language dependent (Lehiste, 1975, 1979, 1980a, 1980b). It may be that the question of the outer bound of such effects is not a meaningful one. However, even if the absolute outer bound of such effects is indeterminate, we can ask what these effects tell us.

Underlying an interest in many of these phenomena is what Monsell and Sternberg (1982) have called the *utterance program hypothesis*. "Certain basic assumptions [about theories of speech production] seem to be widely shared among psychologists, linguists, and other students of speech. One such is the claim, explicit or implicit, that the motor events of an utterance are controlled by the execution of a plan or program--an integrated and relatively detailed description of the utterance (or a large part of it) constructed as a whole before the utterance begins. We term this claim the *utterance program hypothesis*" (p. 1). The utterance program hypothesis has been used in explanation of coarticulation itself and in connection with discussions of related phenomena, such as declination and slips of the tongue. By considering the latter kinds of phenomena first, we can perhaps clarify our discussion of coarticulation theories, which follows.

First, however, we note that many discussions of speech motor plans are circular. An observation is made that speech is normal in one circumstance and abnormal in another. The difference is attributed to the correct or incorrect functioning of a motor plan. For example, our typical student of speech pathology has heard that the articulation difficulty of some populations is due to the failure of a motor plan. The important thing to note is that the invocation of the motor plan adds nothing to the behavioral observation that the population does not speak normally (Kelso & Saltzman, 1982).

A somewhat veiled version of this circular kind of argument is one in which the naked motor plan is given some kind of neuroanatomical or neurophysiological clothing. For an example outside of speech, the control of many kinds of rhythmic activity, such as walking, has been ascribed to the behavior of neural oscillators (Gallistel, 1980), which are not independently observed. While one might not wish to return to the kind of anathema on physiological theorizing dictated by Skinner (1938), it is important to recognize that one of the motivations for his prohibition still holds--there is no explanatory power in the restatement of an observation in different language, even when the language has an independent prestige. Thus, we find out nothing about an aphasic's speech by saying that it is due to the malfunctioning of a particular neural circuit unless we are experimentally prepared to launch a search for the circuit or unless what we know from other sources about a neural circuit of the proposed type allows us to make inferential predictions that we can test about the resultant behavior of aphasics.

A related problem with the metaphor of the motor plan has been raised by Turvey and his associates--it is that the existence of behavioral system activity does not require a single controlling mechanism that lies at a particular level in the nervous system and specifies in detail the properties to be controlled (Turvey, 1977), and, indeed, there are logical problems with the whole idea of a single control center. It may be that some of the characteristics of motor control, which have been attributed to the operation of a plan, are properties of the motor system itself, which emerge as it behaves. Thus, the fact that bees build hexagonal honeycombs does not mean that the bee has a hexagonal floor plan in his central nervous system--rather, the honeycomb may arise in its hexagonal form as a consequence of the interactional properties of the bee and his environment, as the honeycomb is constructed.

Given that we observe carefully these prohibitions on how much we attribute to motor plans, let us return to what we know about the temporal organization of speech. We will talk largely, but not entirely, about precursory effects--that is, the effects that indicate anticipation of speech output before it occurs. While precursory effects tell us nothing about their causes, they tell us something about relevant temporal domains.

An important and much studied phenomenon is the slip of the tongue, in which exchanges occur between elements in speech. A famous example was produced by William Spooner, an English clergyman, who once said, "You've hissed my mystery lectures" in place of "You've missed my history lectures" (Fromkin, 1971). Slips of the tongue are important to a discussion of coarticulation for several reasons. The first, and most important, is that the primary sublexical exchange units seem to be elements very close to the single phonemic segment (Shattuck-Huffnagel, 1983). Apparently, these phonemic segments are correctly produced for their new positions. The existence of such shifts is probably the best evidence we have of the existence of a premotoric terminal stage in the speech production process (MacNeilage, Hutchinson, & Lasater, 1981). Apparently, the units that shift adapt to their new positions--that is, they are correctly coarticulated with their neighbors. Thus, even though we cannot precisely define the phonemic unit in such a way that we can isolate it in the speech stream, slips of the tongue provide some evidence that a phoneme has reality as an action unit. It is interesting to note that although phones participate as action units, single features do not, evidently, appear in exchange error units (Shattuck-Huffnagel & Klatt, 1979).

A final point may be made about slips of the tongue. The sphere over which they occur appears to be of the general length of a breath group. This is roughly the temporal domain of declination and, perhaps, of durational interaction, but it is substantially longer than the temporal extent over which conventional coarticulation spreads.

Another recently fashionable bit of evidence for speech motor planning is the so-called declination phenomenon--the tendency of utterances to decline in fundamental frequency from onset to termination. This is at the utterance level, an analog of the phenomenon studied by Lehiste, and cited earlier, that the onset of the sentence that comes first in a paragraph is higher in fundamental frequency (F0) than the onset of sentences in later positions. Figure 4 is a fairly typical example of sentence declination. Historically, this tendency has been characterized in two ways; as a terminal fall (Lieberman, 1967) and as declination (Maeda, 1975) through the utterance. Again speaking historically, it has been unclear whether the relevant phenomena should be conceived as localized at the end of the sentence, or as distributed throughout. Given that intonation is almost always studied in the context of syntactically complex and phonetically variable contexts, an experimentally clean decision between these alternatives has been difficult, but at least present thinking favors the declination description. That is, the downdrift in F0 appears to run through the sentence, rather than being localized at the end. A related question is whether the mechanism is passive or active. A passive mechanism would be one in which the generalized downdrift is a simple consequence of some physiological given. It might, for example, be a consequence of an uncorrected tendency for subglottal pressure and, hence, F0 to fall throughout the course of an utterance. An alternative would be that the shape of the fundamental frequency contour, regardless of its proximal physiological cause, is a consequence of active planning of the whole utterance. It has been suggested that the latter picture of events is correct because of the utterance length effect--the tendency of F0 to begin at a higher value in longer utterances. In a "speech planning" point of view, a speaker may begin the contour at a higher level in order to come out in the same place.

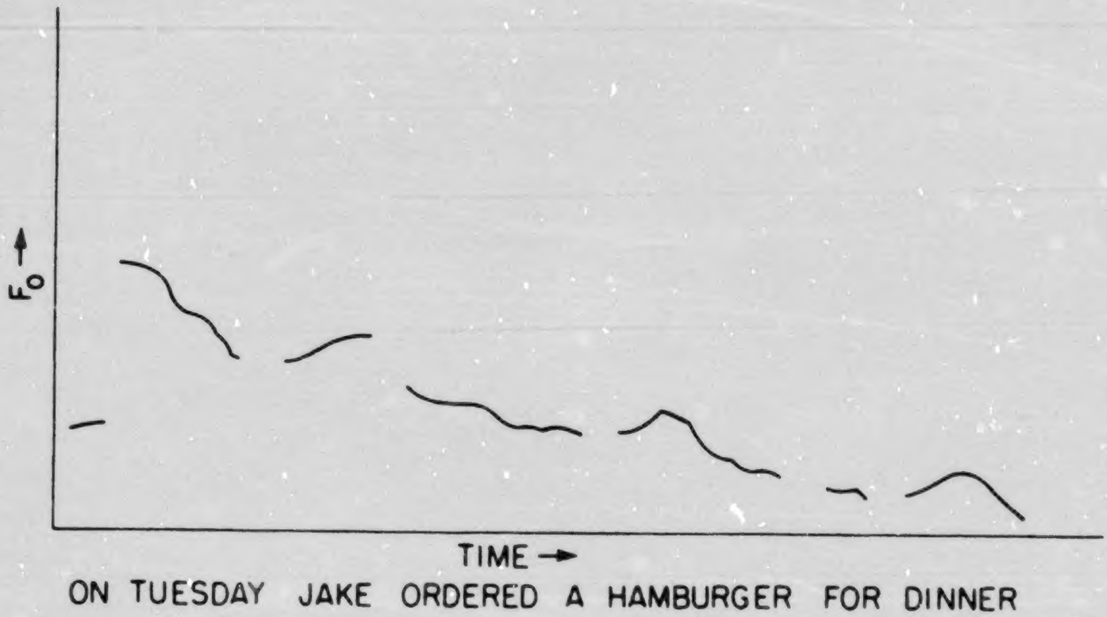


Figure 4. A figure showing declination in a complex "read" sentence. Reproduced from Cooper, W. S., and Sorenson, J., op. cit.

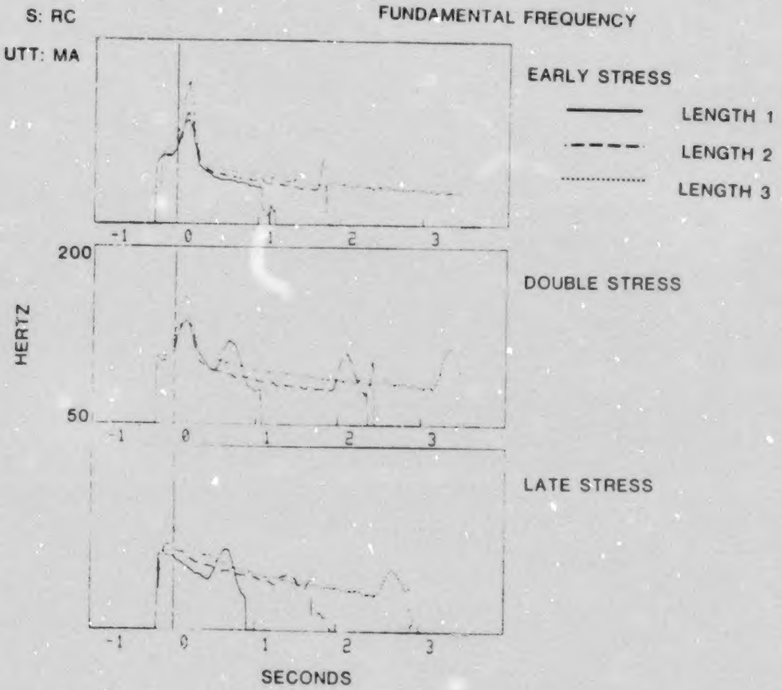


Figure 5. Declination in some sentences with one or two stress peaks. To appear in Gelfer, C., Collier, R., Harris, K. S., and Baer, T. Is declination actively controlled? In I. Titze (Ed.), Vocal fold physiology: Physiology and biophysics of voice. Iowa City: Iowa University Press, in press.

Figure 5 shows F0 contours from a recent experiment by Gelfer, Collier, Harris, and Baer (1983). The contours were produced in reiterant speech--that is, the speaker mimicked himself producing a more complex utterance with the syllable ma. The utterances varied in stress placement and in length. Such structurally simple utterances produce simple fundamental frequency contours made up of one or more peaks. The initial peak varied in amplitude depending on the sentence length, but the effect was very small. From the point of view of speech planning, there were reliable precursive effects, which appear to reflect an overall rough schema for the utterance. However, notice that the whole utterance was not reorganized, depending on its length. Whatever utterance length effects are shown by the declination contour are small and localized. The domain of the effects, however, is the utterance--a domain of about the same size as that for slips of the tongue.

We can say, then, that although speakers may demarcate organizational units of greater length, the longest units over which there is evidence of planning, in the form of precursive effects, is a unit of the general length of a phrase. The examples given here involve slips of the tongue and declination. Similar material could be provided for unit duration. We turn now to conventional coarticulation, which operates over a far smaller temporal domain--on the order of a few speech segments.

Theories of Coarticulation

"Extrinsic Timing" Theories

Since classic theories of coarticulation spring from classic representations of phonological units, such theories almost by necessity attempt to represent coarticulatory phenomena themselves as essentially timeless. In the acoustic real world, no clear boundaries are seen between segments as conventionally defined. Furthermore, acoustic segments are context sensitive; therefore it is necessary to develop some theory that mediates between the acoustic representation and the (presumed) underlying units. Typical examples of such theories are Henke's look-ahead model of coarticulation (Henke, 1966) and Daniloff and Hammarberg's canonical forms model (Daniloff & Hammarberg, 1973). However, other examples of such models can be cited as well; the models as a class were discussed in more detail in a very thorough review several years ago (Kent & Minifie, 1977). Here, we will merely discuss a very well-known example, the Henke model, and refer readers to the review for more detail.

The Henke model assumes that all phonological units can be represented as bundles of features, which occur, in canonical form, as successive units along a time axis. Each phoneme has a specified value, zero, plus, or minus, for each feature. In forming articulatory sequences, the speaker performs an articulatory scanning operation on the phonemes arrayed in a buffer for output. If a feature is unspecified (that is, has a zero value) for several phones preceding the phone for which it is specified, then the feature will be anticipated during the intervening phones; that is, the intervening phones will assume the same feature value as the upcoming one.

Thus, in a sequence of a spread and a rounded vowel separated by nonlabial consonants, the consonants will assume the rounding feature. The test of this thesis has been to ask speakers to produce utterances like once true (Daniloff & Moll, 1968) and then to examine the sequence for the time of the onset of rounding. Using tests of this sort, evidence has been produced that anticipatory coarticulation may spread over as many as four or five segments (Benguerel & Cowan, 1974; Daniloff & Moll, 1968). The model has also been used to explain the early onset of velar lowering in sequences of vowels concluding with nasal consonants. Presumably, in English, vowels are unspecified for nasality; hence, when they precede nasals, they become nasalized.

Success or failure of the model in explaining the data depends on its two assumptions--first, that coarticulatory spread is timeless, and, second, that whatever feature description is made of phones is adequate. For example, Kent and Minifie argue that while the Henke model is assumed to explain the findings of Benguerel and Cowan, it cannot explain the occasional spread of rounding to the end of the preceding spread vowel. One way of giving a "quick fix" to the theory is by relaxing the feature description requirements. Thus, one might assume that the end of a spread vowel is not specified for rounding. A second and more plausible approach is to give up the assumption of timelessness in speech production.

A Temporal Theory

In the light of a "coproduction" theory by Fowler (1980), discussed below, Bell-Berti and Harris (1981) proposed a temporal mode of coarticulation as a substitute for feature-based models. Because Fowler's theory has been somewhat enlarged since it was originally presented, we will discuss the Bell-Berti and Harris view first.

In brief, it was Fowler's thesis that current theories fail to make an appropriate recognition of the temporal dimension in speech production itself. Thus, a theory of anticipatory coarticulation that fails to acknowledge the time course of articulation will fail. She suggested, as an alternative to the view that static elements of vowel and consonant productions are exchanged, that vowel and consonant are coproduced.

A simple model of anticipatory coarticulation, then, makes three propositions (Bell-Berti & Harris, 1981): First, the articulatory period of a segment is longer than its acoustic period; second, for a given articulator, the period of anticipation is temporally independent of preceding phone string number, provided there is no articulatory conflict; and third, that articulatory period may begin at different times for different articulators.

These propositions were tested using electromyographic techniques for anticipatory coarticulation of lip rounding (Bell-Berti & Harris, 1979, 1981, 1982). The test is quite simple. If anticipatory coarticulation is segment based, then its onset will vary with the number of segments; if it is time based, then the duration of anticipatory coarticulation will be independent of the number of segments in a string, provided they do not themselves block coarticulation. Therefore, in order to provide a test, speakers were asked to produce utterances of the type [iC_n u], with a variable number of consonants in intervocalic position. Typical results are shown in Figure 6; the onset of lip-rounding, that is, the duration of the anticipatory period, is independent of the number of anticipatory segments, or of their durations, except for the single voiceless stop condition /itu/.

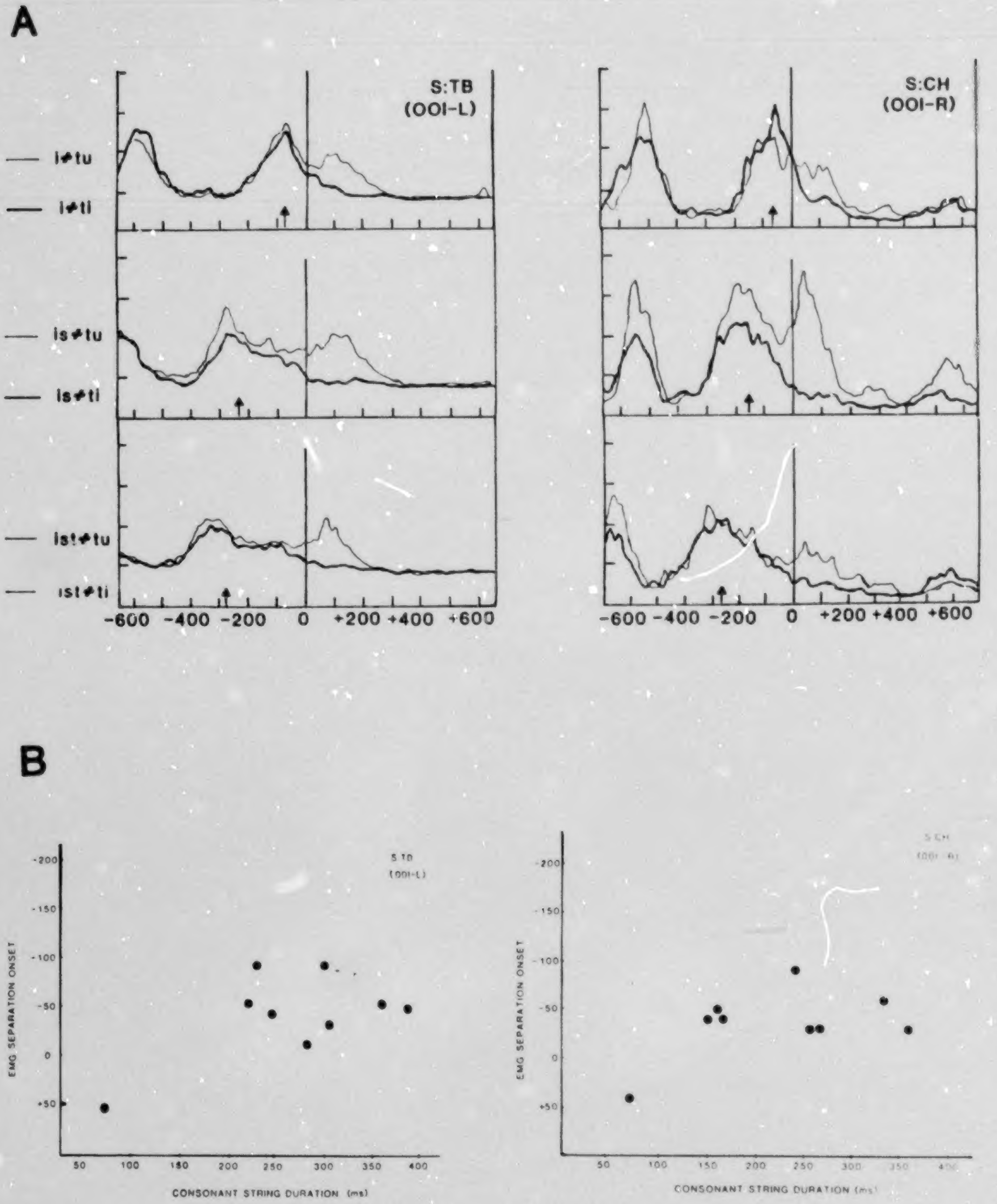


Figure 6. Onset of lip rounding in a series of minimal pairs. Part A shows electromyographic signals from the indicated phone string. Part B shows onset of separation between corresponding pairs. Presented in a paper entitled "Temporal organization of speech units over changes in stress and speaking rate" at the Tenth International Congress of Phonetic Sciences, Utrecht, 1983.

The Bell-Berti and Harris study was repeated, in part, by Sussman and Westbury (1981). They examined anticipatory coarticulation in the strings /kikstu/, /kakstu/, /tiku/, and /taku/. They found different onset times for all four utterances by electromyographic measures, although all differences were not statistically significant. A repeat of the experiment using strain gauge measures found no differences between /kikstu/ and /tiku/. They argued that the failure to find identical onsets for /kikstu/ and /tiku/ or for /kakstu/ and /taku/ argues strongly against time locking of coarticulation to the vowel. They also point out that anticipatory coarticulation is earlier for strings in which the first vowel is /i/ than those for which it is /a/. Their suggestion, in explanation of the latter finding, is that the rounding following /i/ begins earlier because of a necessity to counteract the biomechanical forces that spread the lips for /i/. They support neither the anticipatory scanning model nor the temporally-locked model, although their look-ahead scanner model is segment based.

Both their results and ours point strongly to one experimental issue, noted above; that is, that articulatory constraints are unpredictable from the feature specification of phones. The two vowels /i/ and /a/ are, in feature specifications for rounding, respectively minus and zero. Yet rounding onset time is affected in a manner that is contrary to the feature prediction.

The deviance of the data point for the /itu/ sequence is less conspicuous in the Harris and Bell-Berti data than the /iku/ sequence in the Sussman and Westbury paper, since the latter authors are plotting a two point continuum. On the assumption that the sequences are equivalent in the two experiments, a possible explanation of the deviance of the onset for rounding for single intervocalic stop sequences is provided by Engstrand (1983). He pointed out that relaxation of rounding has been shown to occur in the sequence /utu/ (Gay, 1978; Harris & Bell-Berti, 1983). He suggests that /t/-burst release may be incompatible with a fully rounded lip position. If this is so, then the lips must move rapidly from a fully rounded position, for /u/, to a partly rounded position for the preceding string. In sequences of the form /itu/, full rounding must be suppressed rapidly. For all other sequences (/is tu/; /ist stu/, etc.), while full rounding must end relatively close to the consonant release, partial rounding can end anywhere in the preceding string. The general principle expressed is that production of dentals is incompatible with full rounding and compatible with partial or no rounding. We would, then, expect both the onset and the time course of rounding to be important in a full theory of coarticulation.

Two final experiments on anticipatory lip rounding may be cited--by Lubker (1981) and by Lubker and Gay (1982). The first, by Lubker, gives unequivocal support to the view that the onset of lip rounding varies with the length or duration of the preceding consonant string. The second shows individual differences in the form of the function relating the electromyographic onset of rounding to number of consonants in an intervocalic string. However, this study did not examine consonant string duration but merely consonant number.

In all of the above, we have concentrated on the anticipatory coarticulation of lip rounding. It should be pointed out that there is a similarly detailed, and almost as confused, literature on anticipatory nasalization. Indeed, Al-Bamerni and Bladon (1982) suppose that there may be two forms of anticipatory nasalization--one time locked and one variable. However, this seems a heuristically unsatisfactory solution.

In reading through this account of a series of experiments with their disparate results, the reader should be forgiven for some feeling of bewilderment. It may be worthwhile to consider what we do know. First, it is clear that conventional feature descriptions of phones are not strong enough to predict the details of their articulation, either spatially, that is, in terms of their detailed articulatory topology, or temporally, in terms of when one articulatory gesture begins with respect to another. At present, we are not sure why there are experimental differences among investigators. The only present solution seems to be a more thorough investigation, using simultaneous electromyographic, acoustic, and movement techniques.

Coarticulation and compensatory shortening. Fowler's comments on extrinsic timing theories of speech production have been cited above. However, the theory is far richer and more complex than we have indicated. It was developed, in part, as a means of explaining perceptual isochrony, the phenomenon that syllables perceived as being of more or less equal duration are systematically unequal. Some of its principles form a general theory of production.

Fowler assumes, following Ohman (1966) and Perkell (1969), that vowels and consonants are coproduced so that neighboring segments overlap; i.e., a consonant is produced while a vowel is being produced. The speaker can use such a strategy because vowels and consonants are different kinds of units. Succeeding vowels are produced as slow changes in the position of the tongue body in the mouth. Consonant production is more localized, may involve a partially non-overlapping set of muscles, and is superimposed on the continuous vowel-to-vowel movement of the tongue. Unstressed vowels are presumed not to interrupt the trajectory from one stressed vowel to the next. This model has both spectral and temporal consequences. Let us first consider the temporal consequences.

It has been shown, very often, that the measured duration of a vowel shortens as increasing numbers of consonants are added to it (for a review see Lehiste, 1970). There are backward shortening effects reported as well; that is, a vowel shortens as increasing numbers of consonants precede it (e.g., Lindblom & Rapp, 1973). However, backward shortening (that is, effects of consonants on succeeding vowels) is much the smaller effect. The effects of unstressed vowels on stressed vowels are analogous to the effects of consonants on vowels--for example, the stressed vowel in easy is shorter than in easily. In Fowler's model, the reason for the shortening is the articulatory overlap produced by coproduction.

The same mechanism produces spectral coarticulation. If an unstressed vowel is preceded or followed by a stressed vowel, it should coarticulate with it. Indeed, coarticulatory and shortening effects are but two measures of the same thing and should be highly correlated (Fowler, 1981). Fowler's test of the prediction shows usually significant correlation but some failures in the detailed prediction, apparently due to peculiarities of the particular experimental paradigm.

This theory does not make any predictions about lip rounding, because it is concentrated on the vocal tract manifestations of coproduction, which was the example used by Ohman. It is hard to believe, however, that some parts of the system operate on different principles than others. Furthermore, the model does not cover the well-known shortening effects of consonants on other consonants (Hawkins, 1973). Perhaps its most serious shortcoming, however, is that it does not deal with competing articulation--the circumstance in which the articulators are constricted during consonant production so that free vowel-to-vowel coarticulation cannot take place. For example, Recasens (1983) has shown that in Catalan, vowel-to-vowel spectral coarticulation in vowel-consonant-vowel (VCV) disyllables varies systematically with the extent to which the intervening consonant engages the blade region of the tongue and, consequently, makes coarticulation physically impossible. If Fowler's theory were literally correct, one would expect that differences between VCV sequences in the extent to which they can be coproduced would be accompanied by corresponding differences in the amount of possible compensatory shortening.

Coarticulation and Context Sensitivity

The laboratory investigations discussed above are, perhaps, of interest to the speech pathologist in terms of what they can tell him or her about the practical problems of helping a client to improve a misarticulated sound. What, if anything, have we learned that is relevant?

It is the common observation that certain phonetic environments facilitate correct sound production; for example, Curtis and Hardy (1959), in a now classic paper, showed that some allophones of /r/ are more often correctly produced than others by misarticulating children. As Kent said, "An optimistic interpretation of this contextual facilitation is that some phonetic environments facilitate correct sound production and this facilitation can be exploited to clinical advantage" (Kent, 1982, p. 66). The limits on contextual generalization as a teaching strategy are entirely outside the province of this paper. However, what we can say something about, as a consequence of this brief review, is the task facing the child in learning to talk and the investigator in attempting to specify the contexts that may be relevant subjects of investigation. There are at least two factors that we will need to learn more about:

1. Relative production variability. The first section discussed the insecurity that an observer should feel in making inferences about the articulatory details of production from perceptual judgment. The observer is right, by definition, in judging a child's production to be correct. What he or she cannot do is to infer the articulation from the acoustics, the effects of perceptual factors on his criterion, or the nature of the articulatory error when the speaker is judged to be wrong. Even with respect to the variability of the acoustic signal for a given phonemic percept in a given environment, it is obvious that there is more acoustic production variability in some environments than in others. Some contextual effects may be contextual effects on listener criterion. For example, the formant values for correct stressed vowels are less variable than for unstressed vowels (Summers & Soli, 1982). A more often studied case is /s/, a phone that is notoriously difficult for children and also notoriously subject to contextual inconsistency (Mazza, Schuckers, & Daniloff, 1979). It may be that part of the contextual variability is associated with criterion variability, rather than articulatory variability.

2. Context specification. A lesson to be learned from the literature on coarticulation is that a decision to consider sounds as dividing into allophonic classes leads to balkanization. However, it is questionable whether House's (1981) suggestion that improved transcription may lead to better accounts of context sensitivity will help. A sound can be shown to be differ-

ent in endless ways, depending on factors both within and without the transcriptional record. In truth, we do not know what contexts form natural classes.

It has now been shown repeatedly that children learn phones in words, without uniform generalization across all environments (e.g., Macken, 1980). These types of context sensitivities must have some significance for practical decisions about contexts important in defining a class of phones. On the other hand, certain kinds of context sensitivity are apparently not part of the learning process in children nor are they stored as separately learned patterns in adults. The demonstration that two productions are acoustically or gesturally different does not tell us whether or not the two members form a natural class. It is only careful study of the natural variability of children's articulation, coupled with better assessment of what constitutes motor equivalence and cohesiveness in the adult, that will allow us to make progress in this difficult field.

References

Al-Bamerni, A., & Bladon, A. (1982). One-stage and two-stage temporal patterns of velar coarticulation. Journal of the Acoustical Society of America, 72, S104. (Abstract)

Bell-Berti, F., & Harris, K. S. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. Journal of the Acoustical Society of America, 65, 1268-1270.

Bell-Berti, F., & Harris, K. S. (1981). A temporal model of speech production. Phonetica, 38, 9-20.

Bell-Berti, F., & Harris, K. S. (1982). Temporal patterns of coarticulation: Lip rounding. Journal of the Acoustical Society of America, 71, 449-454.

Benguerel, A.-P., & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. Phonetica, 30, 41-55.

Bernstein, N. A. (1967). The coordination and regulation of movements. London: Pergamon Press.

Bilger, R. C., & Wang, M. D. (1976). Consonant confusions in patients with sensorineural hearing loss. Journal of Speech and Hearing Research, 19, 718-748.

Borden, G., & Harris, K. S. (1980). Speech science primer: Physiology, acoustics, and perception of speech. Baltimore: Williams & Wilkins.

Bruce, R. V. (1973). Alexander Graham Bell and the conquest of solitude. Boston: Little, Brown.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Compton, A. (1976). Generative studies of children's phonological development: Clinical ramifications. In D. Morehead & A. Morehead (Eds.), Normal and deficient child language. Baltimore, MD: University Park Press.

Cooper, W. E., & Sorenson, J. M. (1981). Fundamental frequency in sentence production. New York: Springer-Verlag.

Curtis, J. F., & Hardy, J. C. (1959). A phonetic study of misarticulation of /r/. Journal of Speech and Hearing Research, 2, 244-257.

Daniloff, R. G., & Hammarberg, R. E. (1973). On defining coarticulation. Journal of Phonetics, 1, 239-248.

Daniloff, R. G., & Moll, K. L. (1968). Coarticulation of lip-rounding. Journal of Speech and Hearing Research, 11, 707-721.

Engstrand, O. (1983). Articulatory coordination in selected VCV utterances: A means-end view. RUUL 10 (University of Uppsala, Sweden).

Fischer-Jørgensen, E. (1975). Trends in phonological theory. Copenhagen: Akademisk Forlag.

Fowler, C. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing control. Journal of Phonetics, 8, 113-133.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. Phonetica, 38, 35-50.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. Language, 47, 27-52.

Gallistel, C. R. (1980). The organization of action: A new synthesis. New York: Erlbaum.

Gay, T. (1978). Articulatory units: Segments or syllables. In A. Bell & J. B. Hooper (Eds.), Syllables and segments. Amsterdam: North-Holland.

Gelfer, C., Collier, R., Harris, K. S., & Baer, T. (1983). Is declination actively controlled? In I. Titze (Ed.), Vocal fold physiology: Physiology and biophysics of voice. Iowa City: Iowa University Press.

Harris, K. S., & Bell-Berti, F. (1984). On consonants and syllable boundaries. In L. Raphael, C. Raphael, & M. Valdovinos (Eds.), Language and cognition: Essays in honor of Arthur J. Bronstein. New York: Plenum.

Hawkins, S. (1973). Temporal coordination of consonants in the speech of children: Preliminary data. Journal of Phonetics, 1, 181-217.

Henke, W. L. (1966). Dynamic articulatory model of speech production using computer simulation. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

House, A. S. (1981). Reflections on a double negative: Misarticulation and inconsistency. Journal of Speech and Hearing Research, 27, 98-103.

Jones, D. (1956). An outline of English phonetics (8th ed.). New York: Dutton.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Saltzman, E. L. (1982). Motor control: Which themes do we orchestrate? The Behavioral and Brain Sciences, 5, 554-557.

Kent, R. D. (1982). Contextual facilitation of correct sound production. Language, Speech and Hearing Services in the Schools, 13, 66-76.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. Journal of Phonetics, 5, 115-133.

Ladefoged, P. (1967). Three areas of experimental phonetics. London: Oxford University Press.

Ladefoged, P. (1971). Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

Ladefoged, P., Harshman, R., Goldstein, L., & Rice, L. (1978). Generating vocal tract shapes from formant frequencies. Journal of the Acoustical Society of America, 64, 1027-1035.

Lehiste, I. (1970). Suprasegmentals. Cambridge, MA: MIT Press.

Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. G. Nooteboom (Eds), Structure and process in speech perception. New York: Springer Verlag.

Lehiste, I. (1979). Perception of sentence and paragraph boundaries. In B. Lindblom & S. Ohman (Eds.), Frontiers of speech communication research. New York: Academic Press.

Lehiste, I. (1980a, April). Phonetic characteristics of discourse. Paper presented at the Acoustical Society of Japan.

Lehiste, I. (1980b). Phonetic manifestation of syntactic structure in English. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, *14*, 1-27.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology*, *1*, 301-323.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Lieberman, P. (1967). *Intonation, perception and language*. Research Monograph No. 38. Cambridge: MIT Press.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics* (University of Stockholm), *21*, 1-59.

Lisker, L. (1974). On time and timing in speech. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12, Part 10, pp. 2387-2418). The Hague: Mouton.

Lubker, J. (1981). Temporal aspects of speech production. *Phonetica*, *38*, 51-65.

Lubker, J., & Gay, T. (1982). Anticipatory labial coarticulation: Experimental, biological and linguistic variables. *Journal of the Acoustical Society of America*, *71*, 437-448.

Macken, M. A. (1980). Acquisition of stop systems: A cross linguistic perspective. In G. Yeni-Komshian & J. P. Kavanagh (Eds.), *Child phonology*. New York: Academic Press.

MacNeilage, P. F., Hutchinson, J. A., & Lasater, S. A. (1981). The production of speech: Development and dissolution of motoric and premotoric processes. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Erlbaum.

Maeda, S. (1975). Electromyographic study on intonational attributes. *Quarterly Status Report*, MIT Research Laboratory of Electronics.

Matthews, P. B. C. (1981). Muscle spindles: Their messages and their fusimotor supply. In V. B. Brooks (Ed.), *Handbook of physiology* (Section I, Vol. II, Part I). Bethesda, MD: American Physiological Society.

Mazza, P. L., Schuckers, G. H., & Daniloff, R. G. (1979). Contextual-coarticulatory inconsistency of /s/ misarticulation. *Journal of Phonetics*, *7*, 57-69.

Moll, K. L., Zimmermann, G. H., & Smith, A. (1977). The study of speech production as a human neuromotor system. In M. Sawashima & F. S. Cooper (Eds.), *Dynamic aspects of speech production* (pp. 107-127). Tokyo: University of Tokyo Press,

Monsell, S., & Sternberg, S. (undated). *Speech programming: A critical review, a new experimental approach, and a model of the timing of rapid utterances. Part 1*. Unpublished manuscript.

Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measures. *Journal of the Acoustical Society of America*, *39*, 151-168.

Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, MA: MIT Press.

Perkell, J. (1980). Phonetic features and the physiology of speech production. In B. Butterworth (Ed.), *Language production: Vol. 1, Speech and talk*. London: Academic Press.

Perkell, J. S., & Nelson, W. L. (1982). Articulatory targets and speech motor control: A study of vowel production. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control*. New York: Pergamon.

Pollack, E., & Rees, N. (1972). Disorders of articulation: Some clinical applications of distinctive feature therapy. Journal of Speech and Hearing Disorders, 37, 451-461.

Potter, R. K., Kopp, G. A., & Green, H. G. (1947). Visible speech. New York: Van Nostrand.

Recasens, D. (1983). Coarticulation in Catalan VCV sequences: An articulatory and acoustic study. Unpublished doctoral dissertation, University of Connecticut.

Shattuck-Huffnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. MacNeilage (Ed.), The production of speech. New York: Springer Verlag.

Shattuck-Huffnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. Journal of Verbal Learning and Verbal Behavior, 18, 41-55.

Sherrington, C. S. (1906). The integrative action of the nervous system. London: Constable.

Singh, S. (1978). Distinctive features: A measurement of consonant perception. In S. Singh (Ed.), Measurement procedures in speech, hearing and language (pp. 93-155). Baltimore: University Park Press.

Skinner, B. F. (1938). The behavior of organisms. New York: Appleton.

Stevens, K. N. (1973). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw Hill.

Summers, W. V., & Soli, S. D. (1982). Syllable type influences the acoustic consequences of variations in lexical stress. Journal of the Acoustical Society of America, 71, S113. (Abstract)

Sussman, H. M., & Westbury, J. R. (1981). The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation. Journal of Speech and Hearing Research, 24, 16-24.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum.

CONTEXTUAL EFFECTS ON LINGUAL-MANDIBULAR COORDINATION

Jan Edwards†

Abstract. Coordination between intrinsic and jaw-related components of tongue blade movement during the articulation of the alveolar consonant /t/ was examined across changes in phonetic context. Tongue-jaw interactions included compensatory responses of one articulatory component to a contextual effect on the position of the other articulatory component. A similar reciprocity has been observed in studies that introduced artificial perturbation of jaw position and studies of patterns of token-to-token variability. Thus, the lingual-mandibular complex seems to respond in a similar manner to at least some natural and artificial perturbations.

Several recent models of speech production have posited that speech gestures are accomplished by groupings of articulators that are temporarily marshalled together to achieve a common goal. This sort of functionally-organized goal-oriented behavior has been variously described in the literature as "motor equivalence" (Abbs, 1979), as "coordinative structures" (Kelso, Tuller, & Harris, 1983), and as "functional synergies" (Kelso, Tuller, & Fowler, 1982). All of these models have hypothesized that the lingual-mandibular complex operates as one of these functional synergies during the production of vowels and of alveolar consonants.

Earlier studies of lingual and mandibular activity have revealed several sources of evidence to support this hypothesis. First, it has been observed that jaw height covaries directly with tongue height across vowel categories, although the precise nature of this relationship may vary across subjects and across languages (Bell-Berti, Raphael, Pisoni, & Sawusch, 1979; Wood, 1982). Second, the tongue has been observed to compensate in an utterance-specific way for experimental manipulation of jaw position. The well-known "bite block" experiments provide one example of this type of compensation: The first glottal pulse of a vowel produced with an arbitrarily fixed jaw position is reported to have approximately the same formant frequencies as the corresponding unperturbed vowel (Gay, Lindblom, & Lubker, 1981; Lindblom, Lubker, & Gay, 1979; Lindblom & Sundberg, 1971). In addition, a series of dynamic perturbation studies provide evidence that the lips and tongue can compensate for dynamic as well as static perturbation of jaw position. Folkins and Abbs

---

†Also Graduate Center, City University of New York.

Acknowledgment. This work was supported by NINCDS grant NS-13617 to Haskins Laboratories. I am grateful to Katherine S. Harris, Osamu Fujimura, and Betty Tuller for advice on the analysis and to Mary Beckman and Elliot Saltzman for criticism of earlier versions of this paper. Portions of this research were presented at the 106th meeting of the Acoustical Society of America in San Diego, CA, in November 1983.

(1975) applied a resistive load to the jaw during the closing gesture for a bilabial stop. In all perturbed gestures, bilabial closure was still achieved and compensatory responses were observed in both upper and lower lip displacements. This result has been replicated in a number of experiments by these researchers (Abbs, in press; Abbs & Gracco, 1983) and by others (Kelso, Tuller, V.-Bateson, & Fowler, 1984; V.-Bateson & Kelso, 1984). One of these latter experiments (Kelso et al., 1984) provided additional evidence that this compensatory response is utterance-specific. In that experiment, electromyographic activity of the orbicularis oris inferior (OOI) and the genioglossus posterior (GGP) were monitored during repetitions of /bæb/ and /bæz/. Increased activity of the OOI, but not the GGP, was observed when the jaw was perturbed during the closing gesture for /b/ in /æb/; by contrast, increased activity of the GGP, but not the OOI, was observed when the jaw was perturbed during the closing gesture for /z/ in /æz/.

A third source of evidence comes from observations of unperturbed speech. Hughes and Abbs (1976) had examined lower lip (with the jaw component removed) and jaw positions for three vowels across multiple repetitions of each vowel. They found that a negative correlation between lower lip and jaw position resulted in a relatively invariant lower lip resultant position for each vowel. In a similar study, Honda, Baer, and Alfonso (1982) observed a negative correlation between electromyographic activity of the GGP and jaw height for multiple repetitions of the vowel /i/ in one subject. Furthermore, these authors were able to show that the effect of the observed negative correlations was to reduce variability in first and second formant values for the vowel.

Although these three types of observations are consistent with the notion of functional cooperation within the lingual-mandibular complex, it is unclear what the precise model of functional cooperation is or how these observations are to be related within such a model. The results of the jaw perturbation experiments suggest that the tongue and jaw can interact in a compensatory manner in order to preserve a target articulation. Furthermore, the negative correlation between electromyographic activity of the GGP and jaw height observed by Honda et al. (1982) across multiple repetitions of the vowel /i/ suggest that the tongue and jaw may also interact in a compensatory manner during unperturbed speech, at least in response to token-to-token variability. On the other hand, the fact that jaw and tongue height positively covary across vowel categories may simply mean that both articulators function as independent components of the articulatory feature "vowel height." It is of interest, therefore, to determine whether compensatory interactions of tongue and jaw are observed in response to other influences during unperturbed speech. The coarticulatory context is, of course, one of the major influences on both tongue and jaw positions for a particular segment. The observations cited above suggest that either of two patterns of lingual-mandibular coordination might be observed in the face of context-conditioned variability. First, it is possible that positive covariation between tongue and jaw positions will be observed as a function of the coarticulatory context. Second, it is also possible that a compensatory interaction will be observed between tongue and jaw positions for a particular segment in response to a coarticulatory influence of a neighboring segment. This latter possibility is of particular interest because it would support the claim (Sussman & Westbury, 1981) that there may be active responses to coarticulatory influences and that these active responses cannot be described simply in terms of phonological reorganization (i.e., feature-spreading).

The present experiment was designed to examine the effects of contextual variability on lingual-mandibular coordination during unperturbed speech. Tongue blade and jaw positions for /t/ were analyzed in V_1CV_2 utterances in which the identities of the preceding and following vowels were systematically varied in order to produce systematic variation of articulator positions for the consonant. The data were taken from the existing X-ray microbeam corpus (Miller, 1983). The advantage of this was that it afforded direct observation of tongue position over a relatively large number of repetitions (four per utterance type), at least in comparison to conventional X-ray studies of tongue position during speech. The disadvantage, however, was that the data of only a single subject could be analyzed, given the two criteria that were used to select the utterances for analysis: one, that the phonetic context be comprised of a syllable-initial /t/ preceded by an unstressed but non-reduced vowel and followed by a stressed vowel; and two, that the tongue blade pellet be within 10 mm of the tongue tip.

In order to examine the fine structure of lingual-mandibular coordination, "resultant" movements of the tongue blade (measured in a fixed spatial reference frame) were decomposed into two parts, an intrinsic component and a jaw-related component that reflects the fact that the tongue rests on the jaw. Contextual influences on these components could, in principle, result in any one of three patterns of tongue-jaw interaction. First, it is possible that there is no systematic relationship between the components of resultant tongue blade movement across phonetic contexts. Second, it is possible that the tongue blade and jaw covary with a coarticulatory influence in the same manner as they covary across different vowel heights. In this case, the tongue blade resultant would display as much or more variation in position as its two components across different phonetic contexts. Third, it is possible that the tongue blade and the jaw respond to a coarticulatory influence as they do to an artificially-induced perturbation or to token-to-token variability; that is, one articulator may compensate for a coarticulatory influence on the other articulator in order to preserve an utterance-specific vocal tract shape or acoustic goal, e.g., formation and release of the /t/ closure. In this case, less variation in position would be observed for the tongue blade resultant than for either of its components across different phonetic contexts.

Method

Instrumentation

The X-ray microbeam system at the University of Tokyo (Kiritani, Itoh, & Fujimura, 1975) was used to track the movement of pellets attached to the tongue blade and to a lower front tooth in the x and y dimensions of the mid-sagittal plane. The tongue blade pellet placement for this experiment was approximately 10 mm posterior to the tongue tip. Pellet positions were recorded every 6.8 ms and subsequently synchronized with the simultaneously recorded acoustic speech signal.

Speech Sample and Subject

The utterances examined were six V_1CV_2 types extracted from the following stimulus sentences:

Bea teats it. Ma teats it.
Bea tots it. Ma tots it.
Bea tats it. Ma tats it.

Thus, the intervocalic consonant was always a word-initial /t/, the preceding vowel was a word-final /i/ or /ɑ/, and the following vowel was /i/, /ɑ/, or /æ/. One adult female speaker of American English (Western Louisiana dialect) spoke four tokens of each stimulus sentence. The tokens were produced in randomized order.

Data Processing and Analysis

The axes of the reference frame used to record movements of the tongue blade resultant and jaw were rotated so that one of the rotated axes would correspond to the first principal component of variation for jaw movement. All analyses were performed using this new rotated reference frame aligned with the primary direction of jaw movement.

The simplified model of jaw movement that was used to separate resultant tongue blade movement into its intrinsic and jaw-related components is shown in Figure 1. Jaw movement was modeled as pure rotation about a hinge axis passing through the condyles. Given the relative pellet positions used in the X-ray microbeam data acquisition, it was estimated that about 80% of jaw movement was reflected in resultant tongue blade movement.[1] The mean of the jaw distribution was taken as the reference position for the jaw; intrinsic tongue blade positions were derived on a frame-by-frame basis by subtracting 80% of the difference between the observed jaw position and the jaw mean from the tongue blade resultant position.

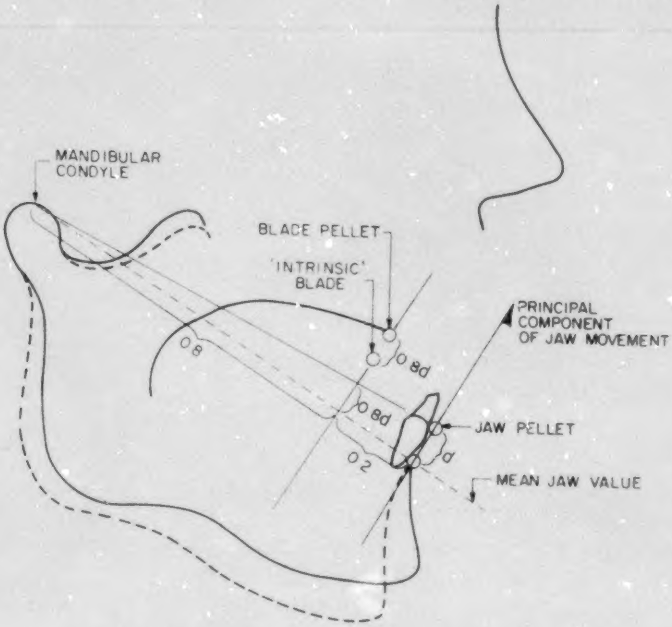


Figure 1. Jaw movement is approximated as simple rotation about a hinge axis passing through the condyles, and coordinates of the tongue blade and jaw are rotated so that the new vertical axis is parallel to the principal component of jaw movement. Since the blade pellet is about 80% of the distance from the condyle to the jaw pellet, 80% of the vertical displacement of the jaw pellet (d) is subtracted from the blade's y-coordinate to get the "intrinsic" blade value.

The y positions in the new coordinate system of the tongue blade resultant, the intrinsic tongue blade, and the jaw were measured at four points in time: acoustic onset of /t/ closure; acoustic release of /t/ closure; peak tongue blade resultant height for /t/; and peak jaw height for /t/. Peak heights were defined as the highest pellet positions occurring at points of zero velocity between the vowel-to-consonant and the consonant-to-vowel transitions. Velocities were derived from the displacement data by the application of a nearly-equal ripple derivative filter (Kaiser & Reed, 1977). Mean displacements of the tongue blade resultant, the intrinsic tongue blade, and the jaw, respectively, for the vowel-to-consonant transitions were 10, 7, and 3 mm for the /it/ gestures and 28, 23, 7 for the /ɑt/ gestures, averaged across final vowels. Mean displacements for the tongue blade resultant, the intrinsic tongue blade, and the jaw, respectively, for the consonant-to-vowel transitions were 5, 2, and 3 mm for the /ti/ gestures; 21, 17, and 5 mm for the /tɑ/ gestures; and 18, 12, and 7 mm for the /tæ/ gestures, averaged across initial vowels. The relative timing of the measured events for most of the utterances was: acoustic closure, blade peak, jaw peak, acoustic release.[2] Figure 2 illustrates the measurement points for one utterance token.

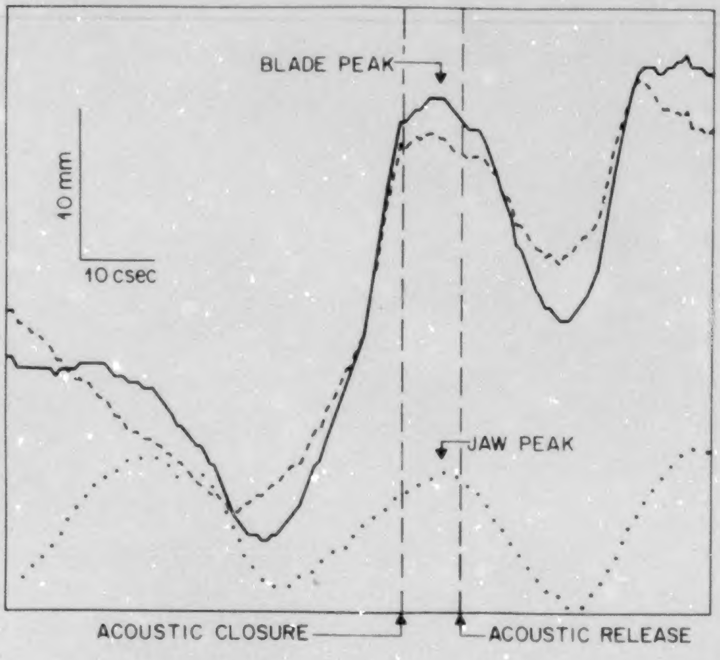


Figure 2. The measurement points (acoustic closure, blade peak, jaw peak, acoustic release) for one utterance token of /ɑtæ/ from the sentence "Ma tats it." The resultant tongue blade is shown in solid lines, the intrinsic tongue blade in dashed lines, and the jaw in dotted lines.

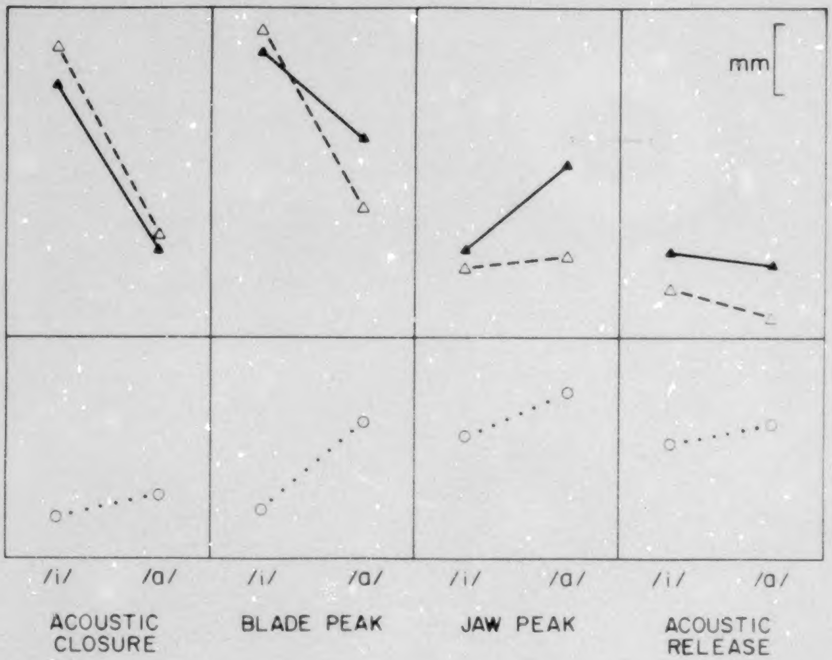


Figure 3. The mean heights of the tongue blade resultant (solid lines), the intrinsic tongue blade (dashed lines), and the jaw (dotted lines) are plotted as a function of the preceding vowel at each measurement point.

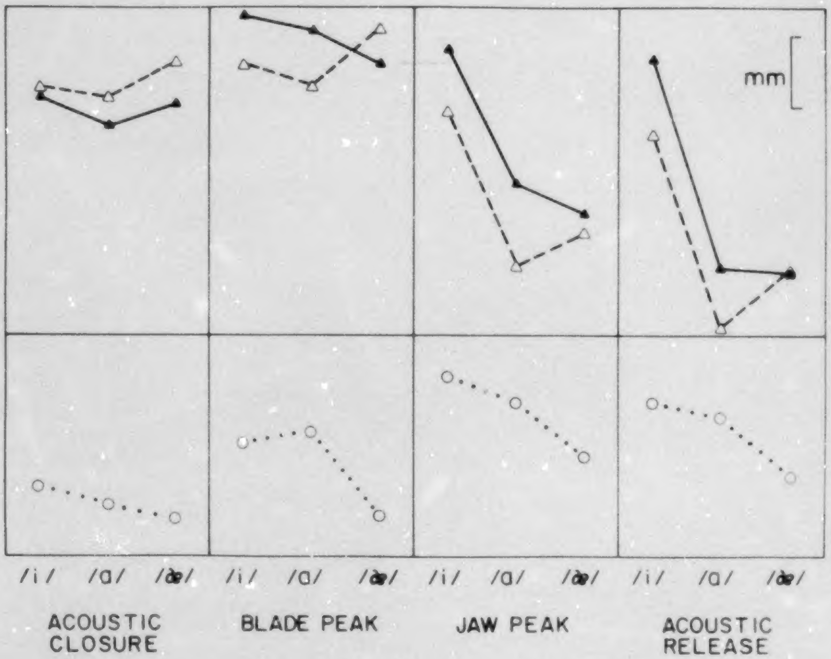


Figure 4. The mean heights of the tongue blade resultant (solid lines), the intrinsic tongue blade (dashed lines), and the jaw (dotted lines) are plotted as a function of the following vowel at each measurement point.

Results

The data are summarized in Figures 3 and 4. Figure 3 shows the mean heights of the tongue blade resultant, the intrinsic tongue blade, and the jaw plotted as a function of the preceding vowel at each measurement point. Figure 4 shows the mean heights of the tongue blade resultant, the intrinsic tongue blade, and the jaw plotted as a function of the following vowel at each measurement point.

In order to assess the magnitude of the effects of the preceding and following vowels, a series of two-way analyses of variance were performed individually for the resultant tongue blade, the intrinsic tongue blade, and the jaw, using the four measurement points. The results of these 16 analyses revealed that the effects of the preceding and following vowels are time-dependent; that is, the main effects of the preceding vowel are significant at acoustic closure ($p < .001$ for the resultant tongue blade and the intrinsic tongue blade) and at blade peak ($p < .001$ for the intrinsic tongue blade and the jaw; $p < .01$ for the resultant tongue blade), but not at acoustic release. Conversely, main effects of the following vowel are significant at acoustic release ($p < .001$ for the resultant tongue blade, the intrinsic tongue blade, and the jaw), but not at acoustic closure. These findings corroborate the results of previous experiments (e.g., Barry & Kuenzel, 1975; Butcher & Weiher, 1976) and support the hypothesis that movement toward the post-consonantal vowel is not initiated until after consonant closure, as was proposed by Gay (1977). One inconsistency with the previous experiments, however, is that one can identify an influence of the preceding vowel at acoustic release by the significant interaction between V_1 and V_2 for the tongue blade resultant. This interaction is displayed in Figure 5; the mean heights of the tongue blade resultant are plotted for each V_1-V_2 combination at this measurement point. An analysis of this interaction revealed that the V2 /æ/ was the sole basis for this significant effect. Because /æ/ was not used in the VCV utterances of the previous experiments, such an effect could not be observed.

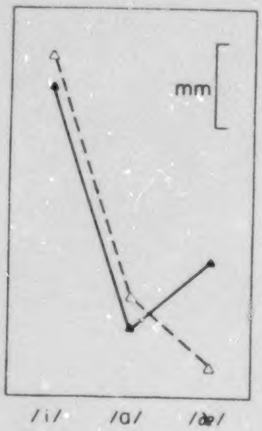


Figure 5. The mean heights of the tongue blade resultant following /ɹ/ (solid lines) and following /i/ (dashed lines) are plotted as a function of the following vowel at acoustic release.

Significant main effects were examined at each measurement point in order to determine if compensatory interactions occurred between articulatory components as a function of phonetic context. An interaction was considered behaviorally salient if it fulfilled two conditions: one, the main effect was statistically significant for both articulatory components; and, two, the direction of the effect was different for the two components for at least one level of that factor. Given these criteria, two instances of compensatory behavior between the components of tongue blade movement were identified: one, at blade peak for carryover influences; and, two, at acoustic release for anticipatory influences. Of course, perfect compensation would yield tongue blade resultant positions that remained invariant across all changes in phonetic context. While the observed compensatory patterns did not produce such an absolute invariance, they did serve to reduce the range of variation in the resultant tongue blade position. Let us consider these two instances of compensation separately.

Carryover coarticulatory influences are illustrated in Figure 3. Consider the second measurement point, blade peak, where a compensatory relationship between jaw and intrinsic tongue blade movements was observed. In this graph, the height of the intrinsic tongue blade varies directly with the height of the preceding vowel: it is 2.5 mm higher after /i/ than after /ɑ/ ($\underline{p} < .001$). The jaw, by contrast, varies inversely with the height of the preceding vowel: it is 1.2 mm lower after /i/ than after /ɑ/ ($\underline{p} < .001$). The net effect of this interaction between the intrinsic tongue blade and the jaw is that the tongue blade resultant displays less variation in position (1.1 mm) as a function of the preceding vowel than does the intrinsic tongue blade.

Anticipatory coarticulatory effects are illustrated in Figure 4. Consider the final measurement point, acoustic release, where another compensatory relationship between intrinsic tongue blade and jaw movements was observed. Post-hoc paired comparisons (Newman-Keuls test) revealed that the means of /ɑ/ and /æ/ are significantly different ($\underline{p} < .05$) for both the intrinsic tongue blade and the jaw. It can be seen in Figure 4, however, that these two means are not significantly different for the tongue blade resultant. This suggests that the tongue and jaw may also interact to compensate for some, but not all, anticipatory influences on /t/ articulation. That is, although the height of the resultant tongue blade is strongly influenced by the degree of constriction for the following vowel (i.e., whether it is high or low), the tongue-jaw interaction serves to reduce the effect of the location of this constriction (i.e., whether it is front or back).

Discussion

The results presented here come from the data of a single speaker producing only four repetitions of six utterance types. Given the ubiquitous intra- and inter-speaker variability that has been found in speech production research, these findings should be interpreted cautiously. Nevertheless, these results suggest that the lingual-mandibular complex responds to some coarticulatory influences in the same manner as it responds to artificially-induced perturbations and to token-to-token variability. That is, the tongue and the jaw may interact in a compensatory fashion, presumably in order to achieve a common goal. Given the data under consideration, it is unclear how to characterize this goal. One possibility is that these tongue-jaw interactions are instances of compensation in order to preserve a target articulation, defined in its most narrow sense. Even though vocal tract occlusion for /t/ is accomplished by the tongue tip, rather than the tongue blade, the position of the tongue blade is constrained in that it cannot fall outside the range of positions that permit tongue tip contact with the hard palate.

Another possibility is that the intrinsic and the jaw-related components of tongue blade resultant position are coordinated in order to decrease the range of variation in the formant transitions during the formation and release of the stop closure. While vocal tract occlusion for /t/ is accomplished by the tongue tip, tongue blade position influences the shape of the cavity behind the occlusion during the final portion of the transitional movement from vowel-to-consonant and during the initial portion of the transitional movement from consonant-to-vowel. A consequence of reducing spatial differences in tongue blade resultant position may be to reduce acoustic variation accordingly. This does not deny the fact that the acoustic transitions vary as a function of the preceding and following vowels. Rather, it suggests that the observed range of variation may be less than what would occur in the absence of these tongue-jaw interactions. This interpretation suggests a line of further research.

Whatever the interpretation, these results provide an example of compensatory inter-articulator coordination in response to contextual influences. Although the data presented here are limited in scope, the results support the hypothesis that observed lingual-mandibular linkages during movement extend beyond a simple mechanical connection between the jaw and the tongue blade. Inter-articulator cooperation, at least for alveolar consonant production, appears to be coordinated to reduce positional variation in resultant tongue blade height generated by the coarticulatory context. The generality of this result, as well as a more detailed description of the conditions under which it is observed, remains to be determined.

References

Abbs, J. H. (1979). Speech motor equivalence: The need for a multi-level control model. Proceedings of the Ninth International Congress of Phonetic Sciences, 2, 318-324.

Abbs, J. H. (in press). Invariance and variability in speech production: A distinction between linguistic intent and its neuromotor implementation. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum.

Abbs, J. H., & Gracco, V. L. (1983). Sensorimotor actions in the control of multimovement speech gestures. Trends in Neurosciences, 6, 393-395.

Barry, W., & Kuenzel, H. (1975). Co-articulatory airflow characteristics of intervocalic voiceless plosives. Journal of Phonetics, 3, 263-282.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., & Sawusch, J. R. (1979). Some relationships between speech production and perception. Phonetica, 36, 373-383.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. Journal of Phonetics, 4, 59-74.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. Journal of Speech and Hearing Research, 19, 207-220.

Gay, T. (1977). Articulatory movements in VCV sequences. Journal of the Acoustical Society of America, 62, 183-193.

Gay, T., Lindblom, B., & Lubker, J. (1981). Production of bite-block vowels: Acoustic equivalence by selective compensation. Journal of the Acoustical Society of America, 69, 802-810.

Gibbs, C. H., & Messerman, T. (1972). Jaw motion during speech. In *Orofacial function: Clinical research in dentistry and speech pathology* (ASHA Reports, No. 7). Washington, DC: American Speech and Hearing Association.

Honda, K., Baer, T., & Alfonso, P. J. (1982). Variability of tongue muscle activity and its implications. *Journal of the Acoustical Society of America*, *72*, S103.

Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica*, *33*, 199-221.

Kaiser, J., & Reed, W. (1977). Data smoothing using low-pass digital filters. *Review of Scientific Instruments*, *48*, 1447-1457.

Kelso, J. A. S., Tuller, B., & Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. *Journal of the Acoustical Society of America*, *72*, S103.

Kelso, J. A. S., Tuller, B., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), *The production of speech*. New York: Springer-Verlag.

Kelso, J. A. S., Tuller, B., V.-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory adaptation to jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 812-832.

Kiritani, S., Itoh, K., & Fujimura, O. (1975). Tongue-pellet tracking by a computer-controlled x-ray microbeam system. *Journal of the Acoustical Society of America*, *57*, 1516-1520.

Lindblom, B., Lubker, J., & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, *7*, 147-161.

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, *50*, 1166-1179.

Miller, J. (1983). The X-ray microbeam database. *Bell Laboratories Technical Memorandum*. Murray Hill, NJ.

Sussman, H. M., & Westbury, J. H. (1981). The effects of antagonistic gestures on temporal and amplitude parameters of anticipatory labial coarticulation. *Journal of Speech and Hearing Research*, *24*, 281-290.

V.-Bateson, E., & Kelso, J. A. S. (1984). Remote and autogenic articulatory adaptation to jaw perturbations during speech: More on functional synergies. *Journal of the Acoustical Society of America*, *75*, S23-S24. (Abstract)

Wood, S. (1982). X-ray and model studies of vowel articulation. *Working Papers* (Lund University, Department of Linguistics), *23*, 1-191.

Footnotes

[1] This model is, of course, physiologically inaccurate in that jaw movement during speech includes both rotation and translation (Gibbs & Messerman, 1972). However, at the level of analysis reported here, the results do not depend on whether the calculation of the jaw component is based on a purely rotational model or on a combined rotation and translation model.

[2] It should be noted that absolute timing (i.e., the durations between each of the measured events) differed systematically as a function of phonetic context. However, a detailed analysis of these differences is beyond the scope of this paper.

THE TIMING OF ARTICULATORY GESTURES: EVIDENCE FOR RELATIONAL INVARIANTS*

Betty Tuller† and J. A. Scott Kelso††

Abstract. In this experiment we examined the effects of changing speaking rate and syllable stress on the space-time structure of articulatory gestures. Lip and jaw movements of four subjects were monitored during production of selected bisyllabic utterances in which stress and rate were orthogonally varied. Analysis of the relative timing of articulator movements revealed that the time of onset of gestures specific to consonant articulation was tightly linked to the timing of gestures specific to the flanking vowels. The temporal stability observed was independent of large variations in displacement, duration, and velocity of individual gestures. The kinematic results are in close agreement with our previously reported EMG findings (Tuller, Kelso, & Harris, 1982a) and together provide evidence for relational invariants in articulation.

A central goal for speech research is to understand the perceptual constancy of a given unit (e.g., feature, phoneme, syllable) in the absence of a unique set of acoustic or articulatory properties. For example, linguistic constraints, such as phonetic context, level of stress, and speaking rate, produce a wide range of articulatory patterns for the same abstract linguistic type. The approach that we adopt here is to ask whether constancies in relational aspects of articulatory patterning (relational invariants) can in fact be observed across these speech-relevant transformations. The present work explores the possibility that the relative timing of articulatory gestures spanning several segments is maintained over suprasegmental variations in stress and speaking rate.

Our interest in the theory that relational invariants (Kelso, 1981) are essential to speech communication is motivated by research from three disparate sources. First, in nonspeech motor skills such as bimanual coordination, handwriting, typewriting, postural control, and locomotion, the relative timing of kinematic or electromyographic events is maintained across scalar changes in rate and force production (see for review, Kelso, Tuller, &

**Journal of the Acoustical Society of America*, 1984, 76, 1030-1036.

†Also Cornell University Medical College.

††Also The University of Connecticut.

Acknowledgment. This work was supported by: NINCDS grant NS-17778 to Cornell University Medical College; NINCDS grant NS-13617 and BRS grant RR-05596 to Haskins Laboratories, and by a grant from the Ariel and Benjamin Lowin Medical Research Foundation. J. A. Scott Kelso was also supported by ONR contract N0014-83-C-0083 to Haskins Laboratories. We would like to thank Carol Fowler, Katherine Harris, Peter MacNeilage, Bruno Repp, Fredericka Bell-Berti, and an anonymous reviewer for comments on earlier versions of the manuscript.

Harris, 1983). For example, as a cat walks faster, the duration of the "step-cycle" of each limb decreases and the propulsive force produced by limb extension increases (Grillner, 1975; Shik & Orlovskii, 1976). However, the timing of activity in the limb extensor muscles is constant relative to the time between successive flexions (Engberg & Lundberg, 1966).

A second source of motivation for examining relational invariants is the demonstration that perception of certain linguistic distinctions depends on the relative (not absolute) durations of acoustic constituents. For example, perception of the voiced/voiceless distinction in medial stop consonants is strongly influenced by the duration of silence (closure) preceding release of the consonant. However, Port (1979) found that the duration of silence necessary to specify that the medial stop consonant was voiceless decreased as speaking rate increased (cf. Miller & Grosjean, 1981; Miller & Liberman, 1979; Pickett & Decker, 1960; Summerfield, 1975).

A third motivation for our approach comes from investigations of speech production. These studies, though few in number, suggest that the relative timing of articulatory kinematics at the segmental and syllabic levels is unaffected by suprasegmental variations (e.g., Kent & Moll, 1975; Kent & Netsell, 1971; Kozhevnikov & Chistovich, 1965; Löfqvist & Yoshioka, 1981).

In an earlier electromyographic study (Tuller, Kelso, & Harris, 1982a), we asked whether stable relative timing across suprasegmental variation is also an appropriate characterization of *intersegmental* speech organization. Specifically, we asked whether the muscle activity underlying production of the vowels and medial consonant in utterances such as /pi#pap/ and /pa#pap/ would maintain any temporal systematicity across rates and stress levels. Our strategy was to define periods of muscle activity corresponding to the interval between successive vowels, and successive consonants. We examined the timing of various aspects of muscle activity for the intervocalic consonant relative to that for the vowel interval, and the timing of muscle events for interconsonantal vowels relative to the consonant interval. Comparing the stability of these various timing relations, we found one very consistent result: The average duration of the interval between onsets of muscle activity for successive vowels was linearly related to the average latency (relative to the first vowel) of medial consonant-related muscle activity.[1] Other possible relationships, such as those based on periods of muscle activity related to production of successive consonants, did not show the same degree of stability.

One shortcoming of our electromyographic experiment (Tuller et al., 1982a) is that we could only examine the stability of relative articulatory timing on the averaged ensemble of tokens. We could not examine whether the relationship also holds when token-to-token variability is allowed because it is not always possible to define onsets and offsets of muscle activity for individual repetition tokens of an utterance (see Baer, Bell-Berti, & Tuller, 1979, for a discussion of temporal measures of individual vs. averaged EMG records). Moreover, the eventual goal is to understand the speech signal as structured by movements of the articulators, but the general form of the relationship between electromyographic signals and kinematic variables is by no means transparent. For these reasons, we performed a similar experiment in which articulator movement trajectories were measured and their relative timing examined.

Method

Subjects

The subjects were three adult females and one adult male. All were native speakers of English. One subject (BT) was aware of the experiment's purpose.

Materials and Procedure

The speech sample included utterances of the form b-vowel-consonant-vowel-b with the medial consonant presented and spoken as the first element of the second syllable. Consonants and vowels were chosen to maximize lip and jaw movement. Thus, the first vowel (V1) was either /a/ or /æ/, the second vowel (V2) was always /a/, and the medial consonant (C) was either /b/, /p/, /w/, or /v/ (e.g., /ba#wab/, /bæ#pab/, etc.). Each utterance was spoken with two stress patterns, with primary stress placed on either the first or second syllable. The subjects read quasi-random lists of these utterances at two self-selected speaking rates--one conversational and the other somewhat faster. Each utterance was embedded in the carrier phrase "It's a _____ again" to reduce the effects of initial and final lengthening and prosodic variations. Three subjects produced twelve repetitions, and one subject (BT) ten repetitions, of the 32 utterance types (8 phonetic strings x 2 rates x 2 stress patterns), for a total of [illegible]472 tokens.

Data Recording

Articulatory movement in the up-down direction was monitored using an optoelectronic device (a modified SELSPOT system). In this system, lightweight, infrared, light-emitting diodes (LEDs) are focused on a photodetector that, with the associated electronics, outputs analog signals corresponding to the x and y position of each LED over time. In this experiment, the LEDs were attached to the subject's upper lip, lower lip, jaw, and nose. In order to minimize head movements during the experiment, a head rest was used and output of the LED attached to the nose was continuously displayed on an oscilloscope placed directly in front of the subject, who was told to keep the display on the zero line.

Acoustic recordings were made simultaneously with the movement tracks and both were computer-analyzed on subsequent playback from FM tape. Acoustic tokens were first excised from the carrier phrase using the PCM system at Haskins Laboratories, then played in random order to four listeners who judged each token's phonetic make-up and stress pattern. Tokens were omitted from further analysis if more than one listener judged the token as having a different stress pattern from the appropriate one or if any phonetic errors were noted. For only one speaker (JE) was it necessary to omit more than two tokens of any given utterance type; the minimum number of utterance tokens for this speaker was seven.

The movement records were computer-sampled at 5-ms intervals. To correct for up-down head movements, output of the nose LED was subtracted (by a computer program) from the output of the LEDs attached to the lips and jaw. Movements of the lower lip were isolated by subtracting movements of the jaw. Velocity records for the jaw, upper lip, lower lip, and lower lip corrected for jaw movement were obtained by software calculation of the first derivative of the position records. For each token, the times at which movements began and ended (indexed by points of zero velocity) were obtained individually for the jaw, the upper lip, and the lower lip corrected for jaw movement.

Results

The main thrust of this study was to examine the relative timing of articulatory movements. In keeping with our earlier work and with various studies of nonspeech motor skills, we chose to define articulatory timing in terms of the phase relations among events in the movement trajectories. This requires delimiting some period of articulatory activity and the latency of occurrence of an articulatory event within the defined period. Over linguistic variations, in this case stress and rate, these intervals will change in their absolute durations. The question is whether they change in a systematically related manner.

Our earlier electromyographic study (Tuller, Kelso, & Harris, 1982a) showed this maximal temporal systematicity when the latency of onset of consonant-associated muscle activity was considered relative to the period between onsets of muscle activity associated with production of successive vowels. We used this result to guide our investigation of articulatory kinematics, although the latencies of gestures associated with vowel events were also examined relative to the interval between gestures associated with successive consonant productions.

Figure 1 shows the acoustic signal and position-time functions for the jaw, upper lip, and lower lip (independent of jaw movement) for one token of /ba#pab/, spoken with primary stress on the second syllable. The figure illustrates the articulatory intervals discussed in the rest of this article. In all cases, the onsets of articulator movements (A through F in Figure 1) were determined empirically from zero crossings in the velocity records of the individual repetitions (not shown in Figure 1). Points labeled A and B are the onsets of jaw lowering associated with production of the first and second vowels, respectively. The interval from A to B is referred to hereafter as the "gestural cycle associated with production of successive vowels" or, more loosely, the "vocalic cycle." Similarly, the intervals from C to D and from E to F are referred to as "gestural cycles associated with production of successive consonants" or "consonantal cycles," indexed by movement onsets of the upper lip and lower lip, respectively. Within the vocalic cycle of each token, we measured the latency of onset of consonant-related movement of the upper and lower lips (i.e., the intervals A-C and A-E). Within the consonant cycle of each token, we determined the latency of onset of jaw lowering associated with vowel articulation (C-B and E-B).

One kinematic measure that is intuitively commensurate with the temporally stable EMG measure is the latency of onset of lower lip raising for producing the medial labial consonant (A-E) relative to the vocalic cycle (the period from the onset of jaw lowering for the first vowel to the onset of jaw lowering for the second vowel (A-B)). These measures are illustrated quantitatively in Figure 2, which shows measurements for one subject's (JE) productions of the utterances /ba#bab/, /ba#pab/, /ba#vab/, and /ba#wab/. Each point on a graph is one token of an utterance type, and the four stress-rate conditions are plotted on a single graph.

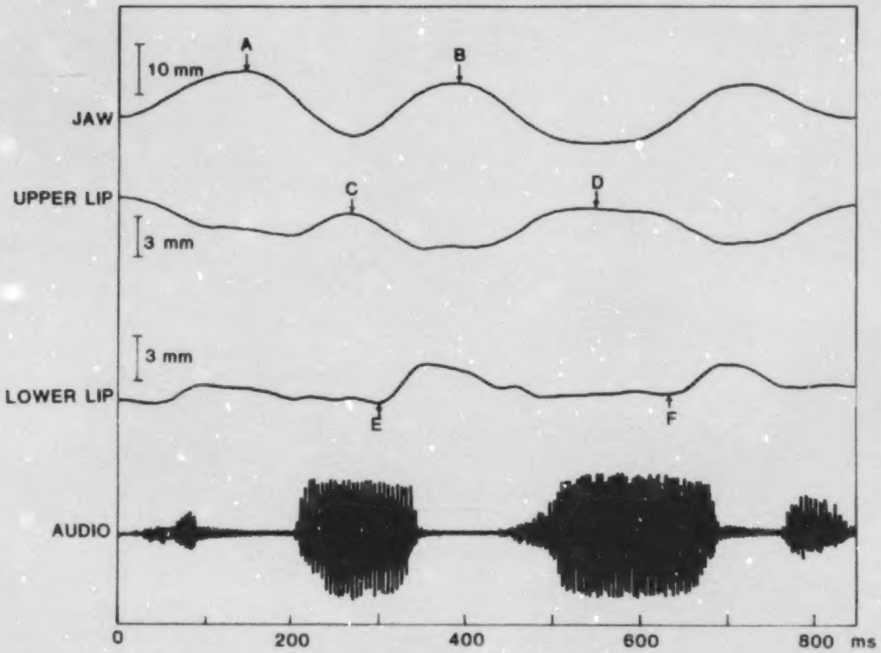


Figure 1. Movements of the jaw, upper lip, and lower lip corrected for jaw movement, and the acoustic signal for one token of /ba#'pab/. Articulator position (the y-axis) is shown as a function of time. Onsets of jaw and lip movements (empirically determined from zero crossings in the velocity records) are indicated (see text for details).

A Pearson product-moment correlation was calculated for each distribution. Obviously, the calculated correlations are very high: .93, .92, .94, and .92. However, the changes that occur are not ratiomorphic; the calculated regression lines (not shown in the figures) do not intercept the y-axis at the origin. Utterances with /æ/ as the first vowel showed essentially identical results, with correlations for this speaker of $\underline{r}$ = .9 and above. Again, the changes were systematic but not ratiomorphic.

Figure 3 also shows the timing of medial consonant articulation relative to the vocalic cycle for the same subject in Figure 2. In this case, however, we have defined the onset of consonant articulation as the onset of the lowering gesture of the upper lip (interval A-C in Figure 1). Utterances with medial /v/ are not included because no systematic upper lip movement was noted. Again, the changes in duration of the two measured intervals are highly correlated for utterances with /a/ as the first vowel (shown in Figure 3) as well as in utterances whose first vowel was /æ/. It can be seen from the figures, however, that correlations within each stress-rate condition tend to be lower than the correlations across conditions, particularly in those conditions whose range is small along one axis.

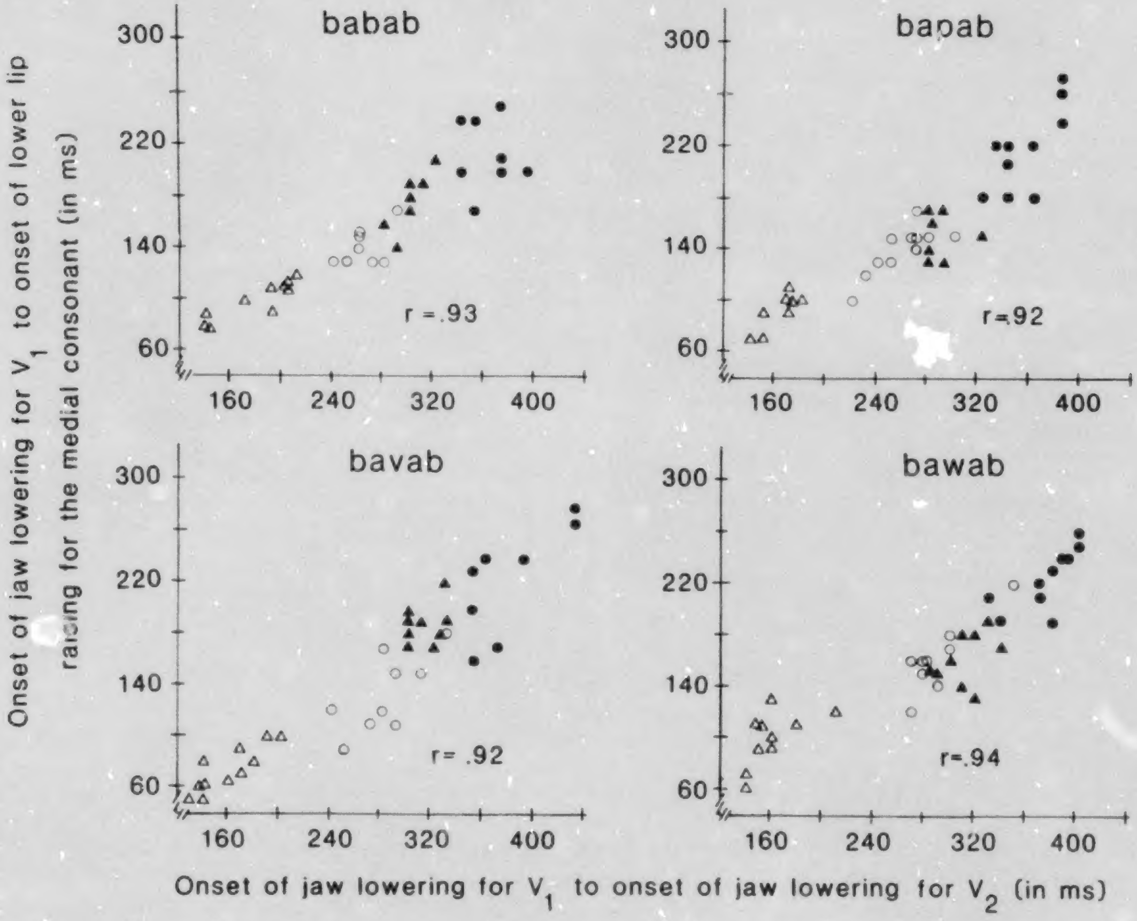


Figure 2. Timing of lower lip raising associated with medial consonant articulation as a function of the vocalic cycle for one subject's (JE) productions of /ba#Cab/ utterances. Filled circles are tokens spoken at a conversational rate with primary stress on the first syllable; open circles are tokens spoken at a conversational rate with stress on the second syllable; filled triangles are spoken at the faster rate with primary stress on the first syllable; open triangles are the faster rate, stress on the second syllable.

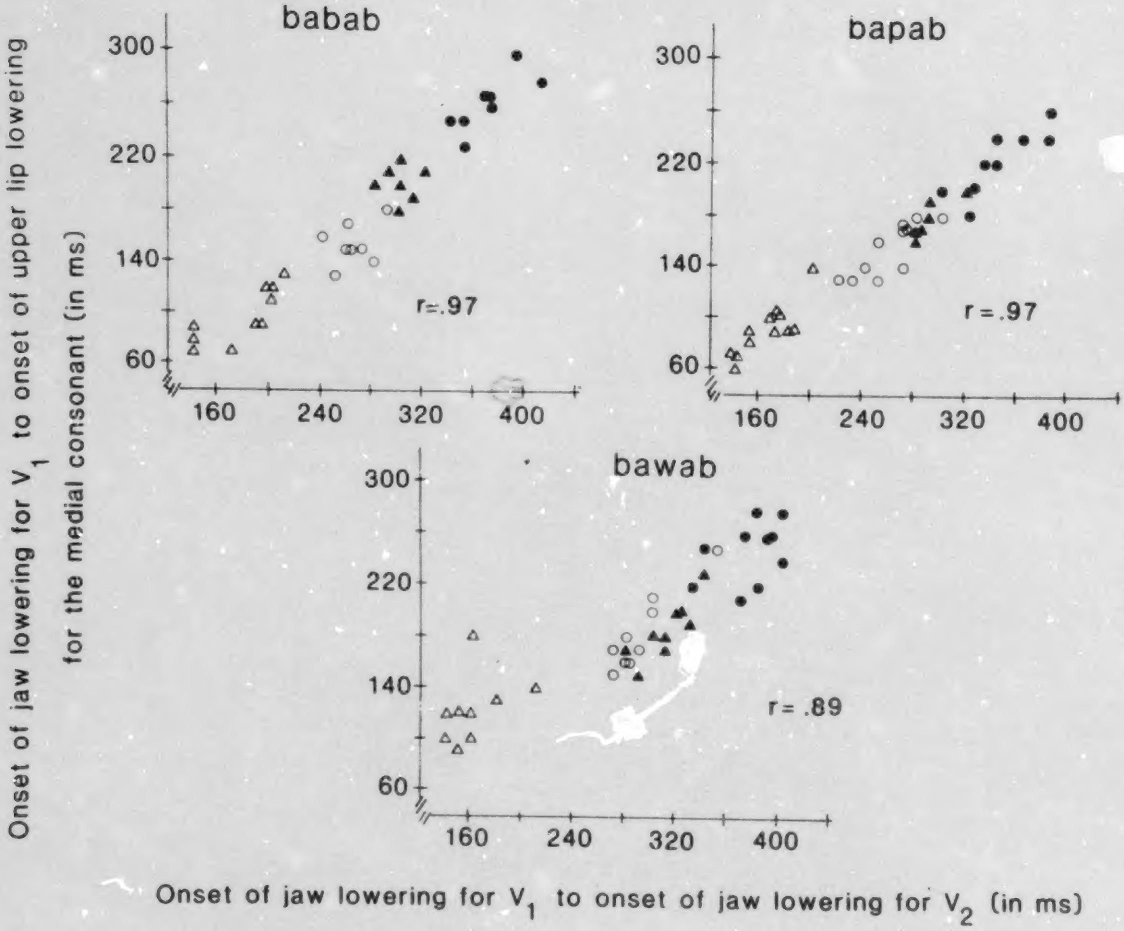


Figure 3. Timing of upper lip lowering associated with medial consonant articulation as a function of the vocalic cycle for one subject's (JE) productions of /ba#Cab/ utterances. Symbols as in Figure 2.

Although Figures 2 and 3 illustrate the data from only a single subject (JE),[2] the three other subjects showed essentially the same pattern. The left half of Table 1 shows the values for all four subjects obtained by correlating the vocalic cycles with the latency of onset of consonant articulation. Correlations obtained when consonant articulation is defined by the raising gesture of the lower lip are shown separately from correlations in which consonant articulation is defined by the lowering gesture of the upper lip. The lowest correlation obtained for any utterance was .84 (accounting for 71% of the variance). Let us underscore that these high correlations occur even though other aspects of the movements, such as their displacement, velocity, and duration, change substantially (Tuller, Harris, & Kelso, 1982). The right half of Table 1 shows the correlations obtained between the within syllable consonantal cycle and the latency of production of the intervening vowel. In Figure 1, these measures correspond to the intervals C-D and C-B for the upper lip and jaw, and E-F and E-B for the lower lip and jaw. The resulting correlations span a wide range of values (from -.02 to .72), clustering in the .2 to .65 range.

One question that arose from this analysis was whether the high correlations obtained between the duration of the vocalic period and the timing of the medial consonant could be a statistical artifact. Most of the durational stretching and shrinking across rate and stress changes occurs in the vowel-related articulator movements. This alone might account for the fact that the correlations between two intervals that both contain the vowel-related movements are higher than the correlations between intervals not containing this common element (cf. Barry, 1983; Tuller, Kelso, & Harris, 1983).

To explore this possibility we determined the correlation coefficients that would occur if consonant gesture latencies occurred at random with respect to gestural periods for successive vowels. To this end, we subtracted the latency (A-C or A-E in Figure 1) from the period (A-B) for all individual tokens of an utterance type. The resulting values (C-B or E-B) were then randomly paired with a different latency value. Adding the members of a pair repairs of values have the same property as our original measure: variability in vowel duration contributes both to period and to latency. We then calculated the correlations between the new pairs. Using Fisher's *r*-to-*z* transform and *t*-tests, we compared the new correlations with the original correlations obtained from the period and latency pairs as measured from the data. Figure 4 shows the difference between the *z*-score for the actual correlation and the *z*-score for the correlation obtained with random pairing of periods and latencies for the 56 comparisons.[3] In all cases, the correlation obtained from the randomly paired periods and latencies was significantly lower (at least at the .05 level) than the correlation of periods and latencies that actually occurred.

A related question is whether our results are due to an overall tempo effect (MacNeilage, in press) and thus do not specifically implicate the gestural cycle for vowels as an important variable in speech motor control. We tested this possibility by examining the interval from the onset of jaw lowering for the second vowel to the onset of upper lip lowering for the final consonant (interval B-D in Figure 1) relative to the interval between jaw lowering for successive vowels (interval A-B). Notice that in this analysis, the defined cycle does not include the relevant consonant-related articulation. Nevertheless, these variables should still be strongly correlated if an over-

---

Table 1

Pearson Product-Moment Correlations for All Four Subjects, Describing Relationships Between Various Periods and Latencies, as Indicated

	Vocalic Cycle				Consonantal Cycle			
	aba[1]	æba[1]	aba[2]	æba[2]	aba[3]	æba[3]	aba[4]	æba[4]
CH	.93	.91	.98	.97	.41	.02	.49	.13
NM	.84	.89	.92	.94	.64	.46	.28	.62
JE	.93	.90	.97	.90	.63	.55	.31	.22
BT	.95	.95	.96	.93	.52	.61	.47	.41
	apa	æpa	apa	æpa	apa	æpa	apa	æpa
CH	.96	.87	.95	.97	-.02	.35	.22	.26
NM	.93	.94	.91	.92	.49	.22	.61	-.02
JE	.92	.94	.97	.89	[illegible]	.29	.36	.64
BT	.97	.96	.96	.93	.71	.31	.46	.21
	awa	æwa	awa	æwa	awa	æwa	awa	æwa
CH	.91	.95	.91	.90	.71	.31	.61	.08
NM	.93	.91	.95	.94	.51	.51	.43	.69
JE	.94	.92	.89	.84	.24	.72	.37	.05
BT	.97	.93	.91	.94	.33	.38	.51	.24
	ava	æva			ava	æva		
CH	.94	.93			.69	.21		
NM	.86	.89			.51	.63		
JE	.92	.95			.46	.52		
BT	.96	.90			.56	.33		

---

[1] Latency = V1 (jaw) to medial C (lower lip); period = V1 to V2 (jaw).
[2] Latency = V1 (jaw) to medial C (upper lip); period = V1 to V2 (jaw).
[3] Latency = C2 (lower lip) to V2 (jaw); period = C2 to C3 (lower lip).
[4] Latency = C2 (upper lip) to V2 (jaw); period = C2 to C3 (upper lip).

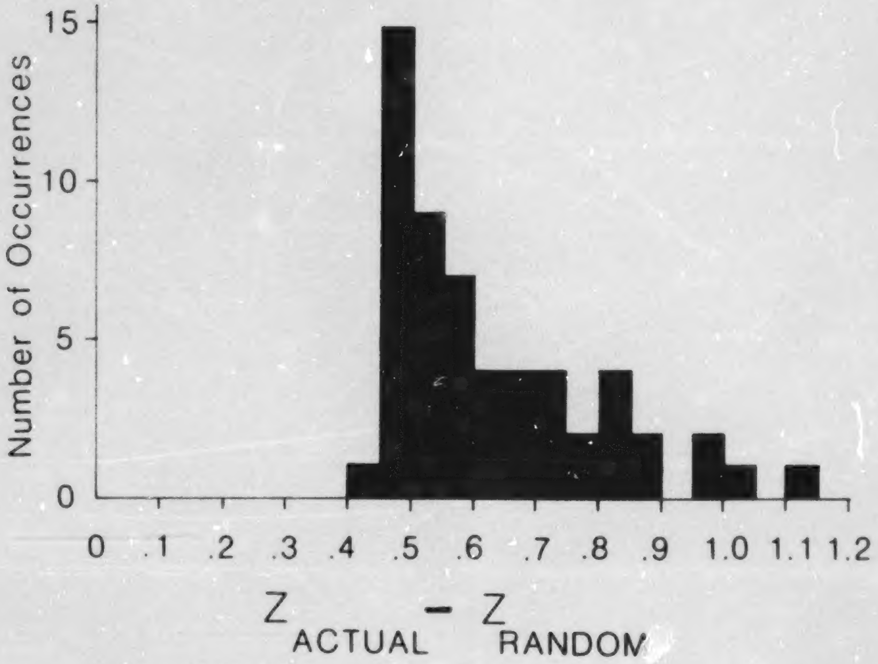


Figure 4. Differences between *z*-scores for the "actual" correlations and *z*-scores for the correlations obtained by random pairing of periods and latencies.

all tempo effect is involved. The resulting linear correlations, however, were extremely weak, ranging from -.6 to .02 across the four speakers, and clustering (83%) in the -.1 to -.4 range. The correlations were generally negative because stressed and unstressed syllables alternate in our data set. Thus, long vowel intervals (utterances with the first syllable stressed) are followed by short lip closing gestures (unstressed, syllable-initial consonants). Taken together with the results of randomly pairing periods and latencies these results indicate that neither variations in vowel duration nor overall speech tempo can account for the systematic relationship between the timing of intervocalic consonant articulation and the period between its flanking vowels.

Another prediction of the stable relative timing of consonant and vowel articulations is that the small changes in duration of consonantal gestures should be correlated with the relatively larger changes in duration of vowel-related gestures. To explore this further, we determined the duration of "vowel-specific movement," defined as the interval from the onset of jaw lowering for the first vowel to the onset of lip movement for the medial consonant (A-C and A-E in Figure 1), and the duration of "consonant-specific movement," defined as the interval from the onset of lip movement for the medial consonant to the onset of jaw lowering for the second vowel (C-B and E-B in Figure 1). We then correlated these measures across stress and rate conditions for each utterance type and, using *t*-tests, determined whether the resulting correlations were significantly greater than zero. In all 56 cases, the durations of consonant and vowel movements (as defined above) were positively correlated (*r*s ranged from .52 to .87, *t*s ranged from 3.55 to 10.29, *p*s < .01).

Although the above analyses examine the commonalities in organization across disyllables with different intervocalic consonants, we expected to observe consonant-related differences predictable from the acoustic-phonetic literature. For example, the period of voicing for a vowel prior to supraglottal occlusion for a voiced stop consonant such as /b/ tends to be longer than voicing for the same vowel before closure for the voiceless stop consonant /p/ (e.g., House, 1961; House & Fairbanks, 1953; Peterson & Lehiste, 1960). For the four speakers in this study, the acoustic duration of the voiced portion of each vowel was measured and ANOVAs computed to test the effect of consonant (/p/ vs. /b/), stress, and speaking rate on vowel-related voicing duration. The acoustic measures were from the first full pitch period after initial consonant release to the first acoustic indication of closure for the medial stop consonant. ANOVA revealed that all four speakers produced significantly longer voicing for vowels before /b/ than before /p/ (*F*s (1,59) ranged from 39.02 to 78.61, *p*s < .001), although for one speaker (CH) this effect was rather small (22 ms mean difference), possibly because the medial consonant was not syllable final.

In light of these results, one might predict that the latency of consonant articulation relative to the preceding vowel (as indexed, for example, by the onset of lower lip raising) would occur later in /b/ than in /p/. Examination of Figures 2 and 3 reveals, perhaps surprisingly, that the range of latencies for the onset of lower lip movement changes only slightly across intervocalic consonants. Although the mean latency values within each stress-rate condition tends to be later for /b/ than for /p/, this small difference does not account for the total measured acoustic difference. The onset of upward jaw movement, however, does migrate with context, being 20 ms to 40 ms earlier in vowel-/p/ than vowel-/b/ utterances.

Another hypothesis is that the period-latency functions might reflect the manner of consonant production. In fact, the calculated regression lines (not shown in the figures) for /v/ and /w/ did tend to have flatter slopes, reflecting earlier articulatory onsets, than the regression lines for /p/ and /b/. However, the ordering of slopes is not identical across subjects. We also evaluated consonant effects on the duration and peak instantaneous velocity of upward movements of the composite lower lip-jaw system. A significant consonant effect was found for both the duration and velocity of lower lip movements for all speakers (*F*s(3,240) ranged from 6.86 to 351.8, *p*s <.001). Scheffé post-hoc comparisons showed that for three of the four speakers, the duration of the lower lip gesture upward was longer for vowel-/v/ and vowel-/w/ transitions than for vowel-/p/ and vowel-/b/ transitions (*p*s <.05). In addition, the peak instantaneous velocity of the composite lower lip-jaw system for all speakers was higher for vowel-/p/, vowel-/b/, and vowel-/v/ gestures than for vowel-/w/ gestures (*p*s <.05). Although the difference in peak instantaneous velocity for vowel-/p/ and vowel-/b/ gestures was just short of significance at the .05 level, all four speakers showed a tendency for vowel-/p/ gestures to have higher velocities than vowel-/b/ gestures (see also Kuehn, 1973).

Discussion

To summarize, in this experiment the timing of movement onset for gestures appropriate to consonants was tightly linked to the timing of movement onsets for vowel-related gestures.* This stability of relative articulatory timing was observed for all utterances examined and was independent of often large variations in duration, displacement, and velocity of individual articulators. Moreover, performance of the one speaker who was aware of the experiment's aim was in all ways similar to the performance of the three naive speakers. These kinematic results are compatible with the earlier EMG findings (Tuller, Kelso, & Harris, 1982a) and together, we feel, provide evidence for relational invariants in articulation. Nevertheless, a few caveats are in order.

First, the measure of movement onset is not meant to be isomorphic with the measure of EMG onset in the earlier experiment. The relationship between parameters of muscle activity and the resulting kinematics has yet to be elucidated in systems far less complex than the vocal apparatus (e.g., Bigland & Lippold, 1954; Cooke, 1980; Wallace & Wright, 1982).

Second, we have chosen to examine the relative timing of onsets of movement trajectories but do not mean to imply that movement onset enjoys privileged status as a directly controlled variable. A good deal of debate in the motor control literature surrounds the question of what variables the nervous system regulates (cf. Stein, 1982, and commentaries). Nevertheless, we feel confident that the timing of onset of articulator movement is highly correlated with whatever kinematic or dynamic aspects of movement are apposite to the nervous system.

A third reason for caution when generalizing these results is that we did not examine the behavior of the most important articulator, namely, the tongue. Although we expressly restricted our corpus to consonants having minimal tongue involvement (so far as we know), any adequate account of speech motor control must include a description of lingual articulation. These data are buttressed, however, by results of a recent, but more limited, parallel experiment that monitored tongue movements of one speaker (Harris, Tuller, & Kelso, 1983; see also Ostry, Keller, & Parush, 1983; Parush, Ostry, & Munhall, 1983).

Fourth, we have only examined phonetically very simple material--the behavior of single consonants between two fairly unreduced vowels, with the intervocalic consonant in syllable-initial position. The description is incomplete in that it does not address the syllable affiliation of the consonant, the number of intervocalic consonants, the role of extremely reduced vowels or schwa, or cases where extensive anticipatory coarticulation is possible.

Despite these limitations, the view that the period between successive vowel gestures is a significant articulatory event and that consonant gestures are timed relative to such periods is supported by the literature on compensatory shortening and coarticulation. For example, it is well known that intervocalic consonants shorten the measured acoustic duration of the surrounding vowels (e.g., Lindblom & Rapp, 1973). This may mean that all aspects of the articulation of vowels are shortened when consonants follow or precede them. Alternatively, it may mean that the consonants and vowels are produced in concert, with the trailing edges of the vowels progressively "overlaid," as it were, by the consonants (Fowler, 1981). In this view, vowel articulations occur continuously throughout the production of consonants and consonant clusters. An articulatory organization of this sort was first proposed by Ohman (1966), to explain the changes in formant transitions of intervocalic consonants as a function of the flanking vowels. Fowler (1977) has elaborated this view by suggesting that the vocalic cycle plays an important organizing role in speech production and perception. More recent articulatory evidence that the influence of both preceding and following vowels is apparent throughout the intervocalic consonant might also be interpreted as indicating a significant vowel-to-vowel articulatory period (Barry & Kuenzel, 1975; Butcher & Weiher, 1976; Gay, 1977; Harris & Bell-Berti, 1984; Sussman, MacNeilage, & Hanson, 1973).

In conclusion, we believe that the data in the study reported here indicate an organizational scheme that speech production shares with many other forms of coordinated activity (see Boylls, 1975; Fowler, Rubin, Remez, & Turvey, 1980; Grillner, 1982; Kelso & Tuller, 1984; Kelso, Tuller, & Harris, 1983; Turvey, Shaw, & Mace, 1978, for reviews), characterized by the temporal stability of movements relative to a cycle and the independence of the relative timing of movements from modulations in displacement or force. In fact, this appears to be one of the main signatures of muscle-joint ensembles when they cooperate to accomplish particular tasks.

References

Baer, T., Bell-Berti, F., & Tuller, B. (1979). On determining EMG onset time. In J. J. Wolf & D. H. Klatt (Eds.), Speech communication papers, 97th Meeting of the Acoustical Society of America, Cambridge, Mass., 1979. New York: Acoustical Society of America.

Barry, W. (1983). Note on interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 826-828.

Barry, W., & Kuenzel, H. (1975). Co-articulatory airflow characteristics of intervocalic voiceless plosives. Journal of Phonetics, 3, 263-282.

Bigland, B., & Lippold, O. C. J. (1954). The relation between force, velocity, and integrated electrical activity in human muscles. Journal of Physiology (London), 123, 214-224.

Boylls, C. C. (1975). A theory of cerebellar function with applications to locomotion; II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. COINS Technical Report (U. Mass., Dept. of Computer and Information Science), 76-1.

Butcher, A., & Weiher, E. (1976). An electropalatographic investigation of coarticulation in VCV sequences. Journal of Phonetics, 4, 59-74.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland.

Engberg, I., & Lundberg, A. (1966). An electromyographic analysis of muscular activity in the hindlimb of the cat during unrestrained locomotion. Acta Physiologica Scandinavica, 75, 614-630.

Fowler, C. A. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. Phonetica, 38, 35-50.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action, In B. Butterworth (Ed.), Language production. New York: Academic Press.

Gay, T. (1977). Articulatory movements in VCV sequences. Journal of the Acoustical Society of America, 62, 183-193.

Gentil, M., Harris, K. S., Horiguchi, S., & Honda, K. (1984). Temporal organization of muscle activity in simple disyllables. Journal of the Acoustical Society of America, 75, S23b (Abstract)

Grillner, S. (1975). Locomotion in vertebrates. Physiological Review, 55, 247-304.

Grillner, S. (1982). Possible analogues in the control of innate motor acts and the production of sound in speech, In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control. New York: Pergamon Press.

Harris, K. S., & Bell-Berti, F. (1984). On consonants and syllable boundaries. In L. Raphael, C. Raphael, & M. Valdovinos (Eds.), Language and cognition: Essays in honor of Arthur J. Bronstein. New York: Plenum Press.

Harris, K. S., Tuller, B., & Kelso, J. A. S. (in press). Temporal invariance in the production of speech. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variance of speech processes. Hillsdale, NJ: Erlbaum.

House, A. S. (1961). On vowel duration in English. Journal of the Acoustical Society of America, 33, 1174-1178.

House, A. S., & Fairbanks, G. (1953). Influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25, 105-121.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance IX. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In M. Gazzaniga (Ed.), Handbook of cognitive neuroscience. New York: Plenum Press.

Kelso, J. A. S., Tuller, B. H., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag.

Kent, R. D., & Moll, K. (1975). Articulatory timing in selected consonant sequences. Brain and Language, 2, 304-323.

Kent, R. D., & Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. Phonetica, 24, 23-44.

Kozhevnikov, V. A., & Chistovich, L. A. (1965). Speech: Articulation and perception. Washington, D.C.: Joint Publications Research Service.

Kuehn, D. P. (1973). A cineradiographic investigation of articulatory velocities. Unpublished doctoral dissertation, University of Iowa.

Lindblom, B., & Rapp, K. (1973). Some temporal regularities of spoken Swedish. Pap. Inst. Ling. (Univ. Stockholm) 21, 1-59.

Lubker, J. (in press). Comment on 'Temporal invariance in the production of speech.' In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variance of speech processes. Hillsdale, NJ: Erlbaum.

Löfqvist, A., & Yoshioka, H. (1981). Interarticulator programming in obstruent production. Phonetica, 38, 21-34.

MacNeilage, P. F. (in press). Comments on 'Temporal invariance in the production of speech.' In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variance of speech processes. Hillsdale, NJ: Erlbaum.

Miller, J. L., & Grosjean, F. (1981). How the components of speaking rate increase influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 208-215.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*, 457-465.

Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, *39*, 151-168.

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 622-636.

Parush, A., Ostry, D. J., & Munhall, K. G. (1983). A kinematic study of lingual coarticulation in VCV sequences. *Journal of the Acoustical Society of America*, *74*, 1115-1125.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Pickett, J. M., & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech*, *3*, 11-17.

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics*, *7*, 45-56.

Shik, M. L., & Orlovskii, G. N. (1976). Neurophysiology of locomotor automatism. *Physiological Review*, *56*, 465-501.

Stein, R. B. (1982). What muscle variable(s) does the nervous system control in limb movements? *The Behavioral and Brain Sciences*, *5*, 535-541.

Summerfield, Q. (1975). *Cues, contexts and complications in the perception of voicing contrasts*. (Belfast, Queen's University, Department of Psychology, Speech Perception No. 4).

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. *Journal of Speech and Hearing Research*, *16*, 397-420.

Tuller, B., Harris, K. S., & Kelso, J. A. S. (1982). Stress and rate effects on articulation. *Journal of the Acoustical Society of America*, *72*, S103. (Abstract)

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982a). Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 460-472.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982b). On the kinematics of articulatory control as a function of stress and rate. *Haskins Laboratories Status Report on Speech Research*, *SR-71/72*, 81-88.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1983). Converging evidence for the role of relative timing in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 829-833.

Turvey, M. T., Shaw, R. E., & Mace, W. (1978). Issues in the theory of action: Degrees of freedom, coordinative structures, and coalitions. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, NJ: Erlbaum.

Wallace, S. A., & Wright, L. (1982). Distance and movement time effects on the timing of agonist and antagonist muscles: A test of the impulse timing theory. *Journal of Motor Behavior*, *14*, 341-352.

Footnotes

[1] This result has since been replicated for speakers of French, using a somewhat more extended phonetic inventory and muscle set (Gentil, Harris, Horiguchi, & Honda, 1984).

[2] Data from a different speaker (CH) are plotted in Tuller, Kelso, and Harris (1982b), and a subset of data from a third speaker (BT) is plotted in Tuller, Kelso, and Harris (1983).

[3] Four subjects X six utterance types X two measures of consonant articulation, plus four subjects X two utterance types with one measure of consonant articulation.

[4] Recent work by Lubker (1983) suggests that for speakers of Swedish, the timing of vowel and consonant movements is constrained as for the English speakers.

ONSET OF VOICING IN STUTTERED AND FLUENT UTTERANCES

Gloria J. Borden,† Thomas Baer, and Mary Kay Kenney††

Abstract. Electroglottographic (EGG) and acoustic waveforms of the first few glottal pulses of voicing were monitored and voice onset time (VOT) measured during an adaptation task performed by stutterers and controls. The fluent utterances of stutterers resembled those of control subjects. After dysfluencies, however, the EGG signal increased gradually, lending physiological support to the technique of "easy onset" of voicing. EGG waveforms also served to help differentiate mild from severe stutterers. Idiosyncratic ritualized laryngeal behavior, sometimes including physiological tremor, was evident in the EGG record.

Physiological studies indicate that initiation of voicing presents particular difficulties for stutterers. Aberrant laryngeal muscle activity (Freeman & Ushijima, 1978) and inappropriate vocal fold positioning (Conture, McCall, & Brewer, 1977) have been found. In addition to abnormally high muscle activity, Freeman and Ushijima found that the usual reciprocity of laryngeal adductor and abductor muscles disappears during stuttering episodes. Conture and his colleagues observed that the vocal folds are fixed during blocks in either a closed or open position. Many methods used to treat stuttering accordingly emphasize "easy onset" of voicing. Van Riper's (1963) technique of altering the preparatory set directed stutterers to start an utterance from a state of rest. Webster's (1974) "Target-based Therapy" and Weiner's (1978) "Vocal Control Therapy" are two of many approaches that direct attention to the gradual onset of voicing. These techniques are supported by numerous studies demonstrating the fluency enhancing effects of conditions (such as choral reading, delayed auditory feedback, metronome-timed speech, and auditory masking) that result in altered phonatory states (Wingate, 1969, presents a review). Also, stuttering episodes were found to become more frequent when changes in voicing were increasingly required (Adams & Reis, 1974).

Even when judged to be fluent, stutterers have been found to be slower than normals in initiating voicing during reaction time experiments (Adams & Hayden, 1976; Cross & Luper, 1979; Starkweather, Hirschman, & Tannenbaum, 1976). Voice onset time (VOT) in CV combinations has also been found to be longer in the perceptually fluent utterances of stutterers than in tokens

---

†Also Temple University, Philadelphia, PA.
††Temple University, Philadelphia, PA.

Acknowledgment. Stutterers were referred by Bernard Stoll and Arlyne Russo, who also tested them for severity. Technical assistance was provided by David Zeichner, Edward Wiley, Richard Sharkany, and Donald Hailey. Figures were drafted by Margo Carter. This study was funded in part by NIH grant NS-13617.

[HASKINS LABORATORIES: Status Report on Speech Research SR-79/80 (1984)]

uttered by normal control subjects (Hillman & Gilbert, 1977), although there have been findings that contradict or qualify the longer VOT results (Metz, Conture, & Caruso, 1979; Watson & Alfonso, 1983). The inconsistency of results in the VOT studies may be due to differences in the degree to which sub-vocal blocks were successfully eliminated from the sample determined to be perceptually fluent. Since the incidence of stuttering episodes is known to be significantly higher at the beginning of a phrase than within it (Bloodstein, 1975), the preparatory "set" does seem to be implicated, but the question remains whether these preliminary adjustments are aberrant in stutterers even when they are fluent. On average, stutterers are slower in their speech than nonstutterers and are also slower in counting on their fingers, but when separated into groups according to severity, a significant difference was limited to the severe stutterers; mild stutterers were not significantly slower than their controls (Borden, 1983).

The phenomenon of adaptation in stuttering, in which the frequency of stuttering episodes is usually reduced in repeated oral readings of the same passage, was exploited in this study to provide examples of fluent and stuttered tokens of an utterance for comparative purposes. In addition, we used the technique of electroglottography (EGG) a useful, noninvasive method of indirectly examining activity of the vocal folds. The recorded EGG signal is the change in impedance across the vocal folds of an imperceptible high frequency current passing between electrodes placed on each side of the thyroid prominence (Fourcin, 1974). To the degree that the vocal folds increase contact with one another, impedance to the transmission of the signal decreases, while glottal opening increases impedance. Thus, vocal fold movements may be inferred from changes in impedance. Investigations comparing the EGG signal with direct filming of the vocal folds have yielded information on landmarks of the EGG waveform and their correlation with glottal opening, closing, and peak contact (Baer, Löfqvist, & McGarr, 1983; Childers, Naik, Larar, Krishnamurthy, & Moore, 1983; Rothenberg, 1981).

The purpose of this experiment (see Footnote 1) was to study the onset of voicing in stutterers and their controls during an adaptation condition for which they repeated 4-digit number series (such as 4253) five times each or until judged fluent. Questions that we had in mind were: What can be inferred about voice initiation from acoustic and EGG analysis

...in stuttered, aborted attempts to voice
...in successful voicing after a block
...in perceptually fluent utterances
...in normal speech of control subjects?

Initiation of voicing was analyzed by examining the acoustic and electroglottographic waveforms of the first few glottal pulses of each of the two number series and by measuring VOT from spectrographic recordings.

Method

Subjects

Eight adult stutterers (seven males and one female) aged 21-48 years were matched by sex, age, and general educational/occupational level with eight normal speakers aged 20-45. Mean age was 33 for the experimental group and 32 for the control group. College students, teachers, blue collar workers, and professionals were represented in both groups. Subjects were bimodally distributed in terms of the severity of their stuttering. Table 1 shows that four of the stutterers were rated as mild and four as severe, according to the Stuttering Severity Index (Riley, 1972), the reading and conversational parts of the Stuttering Interview (Ryan, 1974), and subjective judgments of two speech pathologists.

Table 1

Subjects for the adaptation study and their controls.

	SUBJECT	SEX	AGE	SEVERITY OF STUTTERING
Experimental Group	1. JP	M	48	severe
	2. DE	M	22	severe
	3. DA	M	31	severe
	4. LB	M	44	mild
	5. DL	F	30	severe
	6. MA	M	26	mild
	7. GV	M	41	mild
	8. SL	M	21	mild
			$\bar{X}$= 33	
Control Group	1. FS	M	45	
	2. TS	M	22	
	3. SB	M	30	
	4. EG	M	43	
	5. NM	F	32	
	6. JL	M	29	
	7. AL	M	36	
	8. DR	M	20	
			$\bar{X}$= 32	

Task

Subjects were asked to count aloud from a visual digital display of two different sequences of the digits 2, 3, 4, and 5. The sequences were 3425 and 4253. Subjects were instructed to say each sequence as quickly as possible, without sacrificing accuracy, upon the sound of a response tone. They were told to expect repetitions. Each series appeared five times, the first time 1 s before the signal to respond, and the last four times simultaneous with the signal to respond. If the stutterers were not fluent by the fifth trial of each series, they were instructed to repeat the number series until fluent. All 8 of the control subjects, 3 of the 4 mild stutterers, and 1 of the severe stutterers repeated each series 5 times for a total of 10 utterances from each subject. The remaining mild stutterer and three of the four severe stutterers repeated each series (14,10), (14,10), (10,24), and (11,10) times, respectively. One severe stutterer never fully adapted to the 4253 sequence after 24 trials.

Instrumentation

The program presenting the test sequences was run on a microcomputer (Integrated Computer Systems). For each sequence, a visual warning signal was followed by a variable interval (300, 400, or 500 ms), after which the 4-digit display appeared. The tone signaling the subject to respond was delayed 1 s after the first display of each series and was simultaneous with the display for the repetitions. Presentation of each display was experimenter-controlled to allow for subject differences in response time.

An electroglottograph (F-J Electronics ApS) recorded rapid changes in impedance by high pass filtering (25 Hz-10 kHz) the overall changes in impedance of a signal transmitted across the larynx at the level of the vocal folds. The onset of these rapid oscillations was abrupt and unambiguous and served to signal the onset of voicing during the adaptation task. The acoustic pressure wave was simultaneously recorded through a microphone placed approximately 1 foot from each speaker. Lip/jaw movement was recorded from a small LED, attached to the lower lip, that was exposed to an opto-electronic tracking system; respiratory movements were recorded by a semi-hemispheric pneumograph. The respiratory and lip/jaw recordings were not analyzed in detail for this report (see Note 1).

Analysis of the Data

Visicorder graphs of the physiological and acoustic signals recorded on FM tape were produced for each subject. The adaptation trial recordings were inspected for any sign of dysfluency, such as abnormal fluctuations in laryngeal impedance. The trials were then digitized from the analog tape for further editing. The experimenters inspected each set of trials on a computer monitor using a 100-ms time frame to magnify the first few periods of rapid vibrations of the vocal folds, enabling a more detailed examination of the electroglottographic and acoustic waveforms. Hard copies were made of the first dysfluent and last fluent utterance for each series in the sample collected from stutterers and of the first and last trial from each subject who did not stutter.

In addition to the waveform recordings, sound spectrograms were produced for all utterances during the adaptation series. A total of 223 spectrograms were generated to measure VOT of the utterance two /tu/. VOT was measured from the onset of the burst for /t/ to the first glottal pulse for /u/. If the utterance was stuttererd by repetition of /t/, the measure was taken from the last burst to vowel onset. Measures in millimeters were converted to milliseconds and averaged for each subject and across groups corresponding to (1) utterances of control subjects, (2) fluent utterances of stutterers, and (3) stuttererd utterances. Speech rate was measured for the first and last fluent sample for each speaker, yielding four measures (two number series) that were then averaged. Measures were taken from the onset of voicing for the first syllable to voice offset for the last syllable, thus eliminating the sometimes ambiguous onset of the initial consonant. The measures in millimeters were converted to milliseconds and divided by four for an average time for each syllable. This time divided into 1000 ms yielded an average syllable/second speech rate.

Analysis of the EGG and acoustic waveforms at voice onset was qualitative. Quantitative measures of VOT differences between stutterers and controls were averaged across fluent utterances and the standard deviations computed. Spearman's rho correlation was used to test the relationship between VOT and speech rate.

Results

Electroglottographic and Acoustic Waveforms

Control subjects. The patterns of change in laryngeal impedance recorded by the electroglottograph looked similar for all control subjects. Figure 1 represents the EGG and acoustic waveforms of a male voice initiating /ɔr/ in the word four. The polarity of the signal for this analysis is set so that upward deflection indicates the decreased impedance that accompanies increased vocal fold contact, and downward deflection indicates the increased impedance accompanying decreased vocal fold contact. Normally, vocal fold contact increases more abruptly (a) than it decreases (b). There is a relatively stable open phase (c). The EGG envelope grows rapidly in amplitude (d) relative to the typical acoustic waveform for a vowel after /f/, a waveform that is more gradual in buildup of the envelope (e). In previous studies, direct viewing of vocal fold vibration simultaneous with EGG recordings has established these landmarks of the impedance signal (Baer et al., 1983; Childers et al., 1983; Rothenberg, 1981). It is difficult to determine the moment of glottal opening as the folds peel apart during the downward slope of the signal, although sometimes there is a "shoulder" in the downward slope that corresponds with the appearance of a glottal aperture. Peak EGG is fairly reliable, however, as an indication of maximum vocal fold contact, although it does not necessarily indicate complete glottal closure. Occasionally one sees a cycle of impedance that does not result in an acoustic pulse. This may reflect some prevoicing laryngeal adjustment.

Stutterers when fluent. The first finding from inspection of the EGG waveforms of stutterers during fluent utterances was that the waveforms looked normal, with abrupt closing, gradual opening, a relatively stable open phase, and a rapid buildup of the EGG envelope. Figure 2 shows the waveforms from a male stutterer (severe) and his control and a female stutterer (severe) and her control. All four samples are from the final trial of the series 4253, showing onset of voicing in the word four. There is no obvious difference in EGG and acoustic waveforms of stutterers when they are fluent and those of normal speakers.

Stutterers when dysfluent. The second observation from the data on stutterers was that when the stutterers (whether mild or severe) were dysfluent (six of the eight subjects), voice initiation after a block was characterized by a gradual instead of abrupt buildup of the EGG signal in all but one of the subjects. Table 2 indicates the features observed. Two of the mild

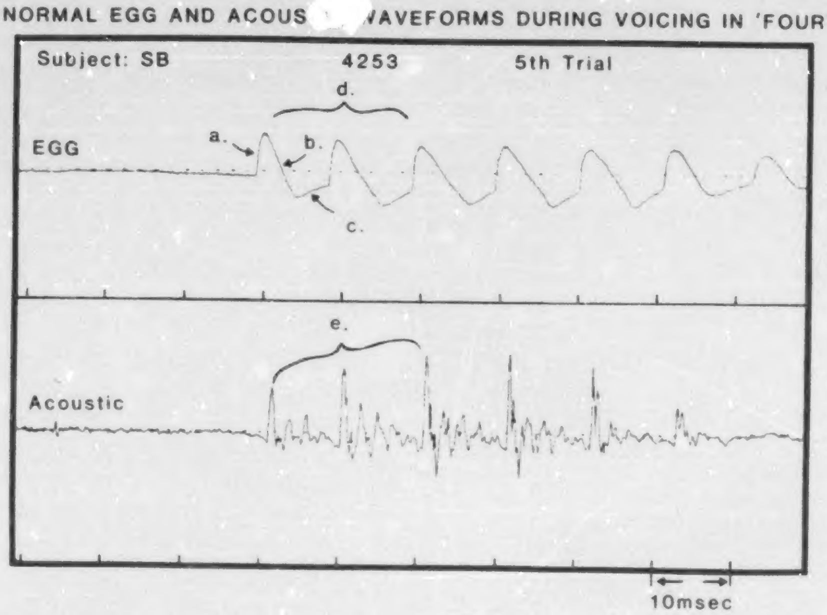


Figure 1. Electroglottographic (EGG) and acoustic records at voice onset in the utterance "four" by a normal subject. The EGG waveform is displayed with upward deflection indicating decreasing impedance. The EGG waveform is characterized by steep rise (a) in 'vocal fold contact' followed by slower 'opening' (b) and 'open phase' (c). Amplitude of the first EGG pulses builds rapidly (d), compared with the acoustic waveform (e).

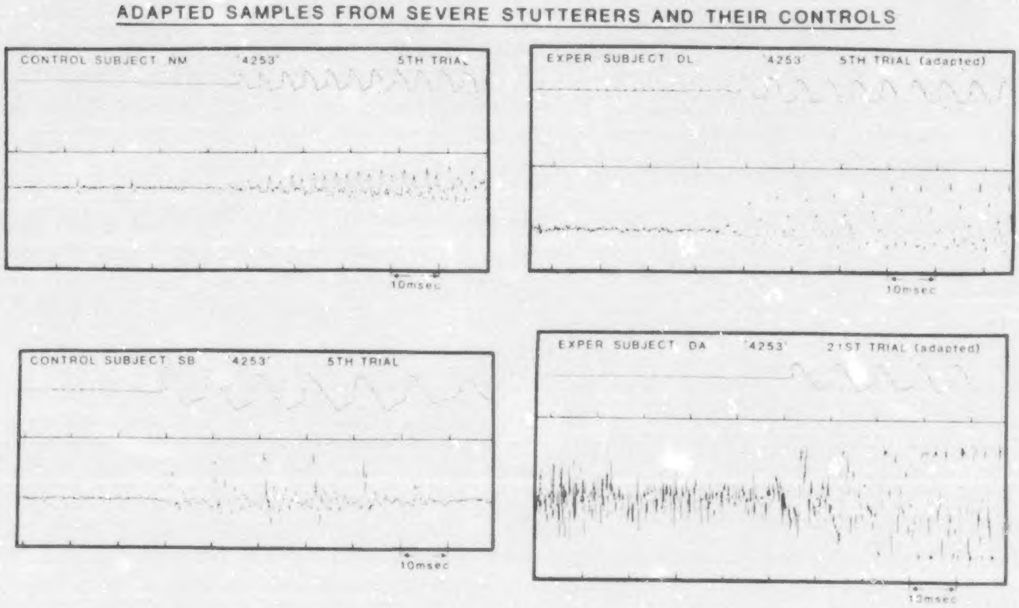


Figure 2. Electroglottographic and acoustic waveforms of normal speakers on the left and of stutterers, when fluent, on the right. The top pair is for females; the bottom pair is for males. Stutterers, when fluent, produce EGG waveforms that build rapidly in amplitude like those of control subjects.

Table 2

Summary of characteristics of the electroglottographic waveforms. All subjects showed rapidly increasing vocal fold contact during each cycle (see Figure 1.a). Thus, this factor did not distinguish mild from severe stutterers. After a block, severe stutterers tended to show more abruptly decreasing vocal fold contact (Figure 1.b) and less stable open phase (Figure 1.c) during the vibratory cycle than mild stutterers, although the normally gradual decrease in contact and open phase were restored when subjects were adapted. The normally abrupt envelope of the EGG signal (Figure 1.d) was not present after stuttering but reappeared when adapted. The occasional pre-voicing EGG cycle occurred for a few stutterers and a few controls and did not serve to distinguish one group from another.

CHARACTERISTICS OF THE ELECTROGLOTTOGRAPHIC WAVEFORMS

		Rapid increase in VF contact	Gradual decrease in VF contact	Stable Open Phase	Abrupt Envelope	Pre-voiced adjustment
CONTROLS		△	△	△	△	
		△	△	△	△	△
		△	△	△	△	
		△	△	△	△	△
		△	△	△	△	
		△	△	△	△	
		△	△	△	△	
		△	△	△	△	△
STUTTERERS MILD	MA	△	△	△	△	
	GV	△	△	△	△	△
	SL	△	△	△	−△^{A}	
	LB	△	△	−△^{A}	−△^{A}	
STUTTERERS SEVERE	DE	△	△	△	△	
	DA	△	−△^{A}	△	−△^{A}	△
	DL	△	−△^{A}	−△^{A}	−△^{A}	△
	JP	△	−△^{A}	−△^{A}	−△^{A}	
		Not distinctive	Severe differed from mild when dysfluent.			Not distinctive

△ present

−△^{A} not present when stuttered, present when adapted stutterers evidenced gradual buildup of the EGG signal after a block until adapted. In other respects the waveforms resembled those of control subjects although the open phase for LB was brief. Severe stutterers when dysfluent, however, differed from normal in several respects: a steeper decrease in vocal fold contact, a less stable or prolonged open phase, and a more gradual buildup of the EGG signal. These differences also disappeared upon adaptation. One of the severe stutterers initiated voicing with normal looking EGG whether dysfluent or fluent. The consonant/vowel ratio was reversed in duration, however. During a dysfluent 3425, silence and consonant noise lasted 400 ms while voicing lasted 200 ms, in contrast with the fluent sample in which the ratio reversed to 1:2 with pause and consonant time 200 ms and voicing 400 ms. The rest of the stutterers evidenced gradual buildup of EGG amplitude to initiate voicing after a block (Figure 3).

This gradual rise in EGG amplitude is a physiological index of "easy onset of voicing." It is a more reliable indicator than the acoustic waveform, because the sound is often graded in rise time due to an increase in front cavity opening of the vocal tract and perhaps an increase in volume velocity from subglottal air pressure. For an utterance such as four, the acoustic waveform typically shows a graded envelope as the oral constriction for the /f/ opens for the vowel. Normally, as we have seen, the EGG waveform is abrupt in the rise time of its envelope, indicating that speakers position their folds for voicing (not necessarily completely adducted) before the aerodynamic forces act upon the folds to set them into vibration. The slow rise time in EGG shown by two of the mild and three of the severe stutterers is abnormal and adaptive. It is a strategy that stutterers apparently use to initiate voicing when they are experiencing difficulty. The strong indication is that under these circumstances the aerodynamic forces are brought into play during a gradual posturing of the vocal folds for voicing, resulting in the slow buildup of the EGG envelope seen in Figure 3. Furthermore, once the stutterers are adapted or "fluent," the EGG envelope is abrupt like that of the control subjects. This style of voice initiation does not seem to be used routinely by stutterers but rather as a method for breaking the block.

Although the phenomenon of gradual EGG buildup was evident for both mild and severe stutterers, two characteristics of the EGG waveforms were more common among severe stutterers. Both the normally gradual decline in the signal corresponding to gradual decrease in vocal fold contact as the folds peel apart and the normally stable open phase are less prominent in the stuttered trials of the adaptation task. Figure 4 shows this change. The somewhat steeper decline in the EGG signal and the brief open phase before they snap closed again as seen in the top part of the figure indicate that the folds in the stutterer were more rigid than normal. Additional evidence of a change in stiffness is the corresponding decline in fundamental frequency of the waveforms when adapted. The vibration initiated after the block was 170 Hz compared with 114 Hz upon adaptation. The bottom part of the figure shows the EGG activity for the control subject.

Another observation is the existence of highly ritualized "break-the-block" behavior. One severe stutterer in our sample demonstrated a 3-stage laryngeal maneuver to initiate voicing that looked similar across different utterances. Figure 5 shows the EGG patterns that accompanied the block and final breaking of the block for the utterances three [θri] in 3425 and four [fɔr] in 4253. When adapted, this subject had an f_0 of 114 Hz for the onset of voicing in both utterances, but the first part of the 3-stage ritual

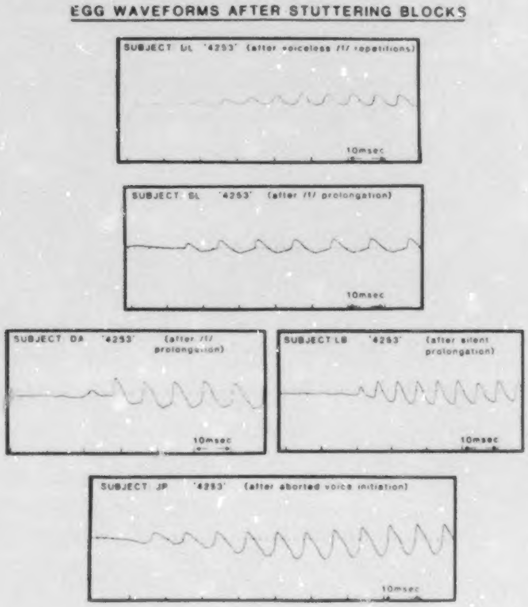


Figure 3. Records from five subjects showing EGG activity following a stuttering block. The more gradual build-up of the EGG envelope is characteristic of most of the stutterers when they initiate voicing after a block.

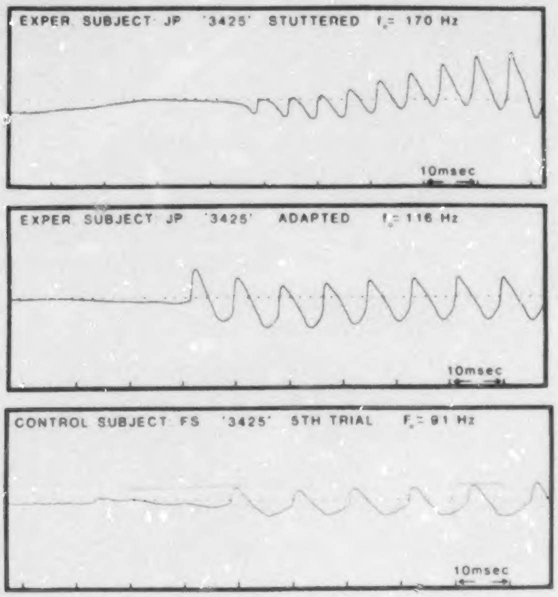


Figure 4. Voice initiation (onset of [ri] in three) in a severe stutterer and his control. Top waveform is the EGG after a stuttering block with the characteristic slow rise time. The relatively steep decrease in vocal fold contact and brief open phase of the waveform is accompanied by high fundamental frequency. The middle waveform is the same utterance adapted to fluency with slightly longer open phase, normal rise time of the first few pulses, and a lower fundamental frequency. The waveform at the bottom of the figure is the EGG signal for the same utterance by the control subject.

used to break the block showed a much higher fundamental frequency. In the trials shown in the figure, the first stage had an f_0 of 170 Hz. It can also be seen that as the EGG signal shows larger impedance changes, the corresponding acoustic signal is gradually lowered in f_0 and finally aborted. The second stage is characterized by breathy low frequency vibrations whose acoustic output is again choked off as the impedance changes widen in their excursions. The third stage is always successful in that voicing is initiated and maintained, although it is abnormally graded in its EGG envelope in contrast to the adapted sample seen in the middle part of Figure 4. Except for the graded EGG seen in the final stage, the rest of the break the block strategy seems maladaptive, as voicing failed to be maintained.

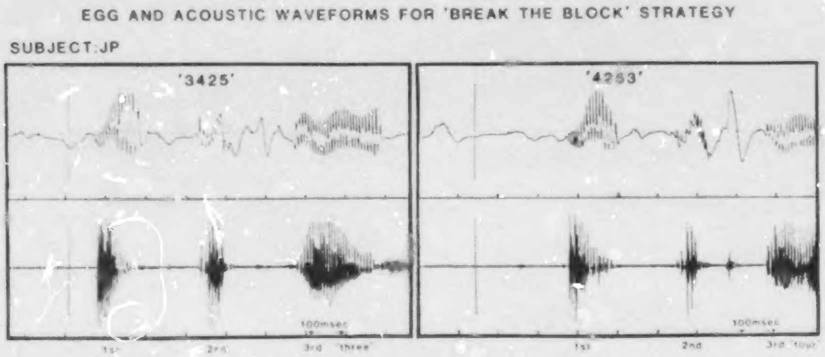


Figure 5. EGG (top) and acoustic (bottom) records associated with two stuttering blocks for one subject. This idiosyncratic and ritualized strategy for initiating voicing after a block is similar despite differences in utterance. This figure shows one second of time, but can be compared with the same subject in Figure 4, which shows 100 ms of voice initiation for 3425.

The final observation from the EGG and acoustic data was the existence of a physiological tremor that shows up on the EGG signal during voiceless blocks. The laryngeal tremor is often phase-locked with an observable tremor in the lower lip. These tremors were observed in two of the severe stutterers. The subject (DA) represented on the top part of Figure 6 had a 9-Hz tremor and the subject (DL) on the bottom had a 7-Hz tremor. These correspond with the lip tremors of 7-10 Hz that Fibiger (1971) recorded by EMG from the facial muscles of stutterers. Physiological tremor has been linked to heightened stretch reflex due to increased gamma motoneuron activity (Lippold, 1971). The data from the second subject shows the 7-Hz lip tremor superim-

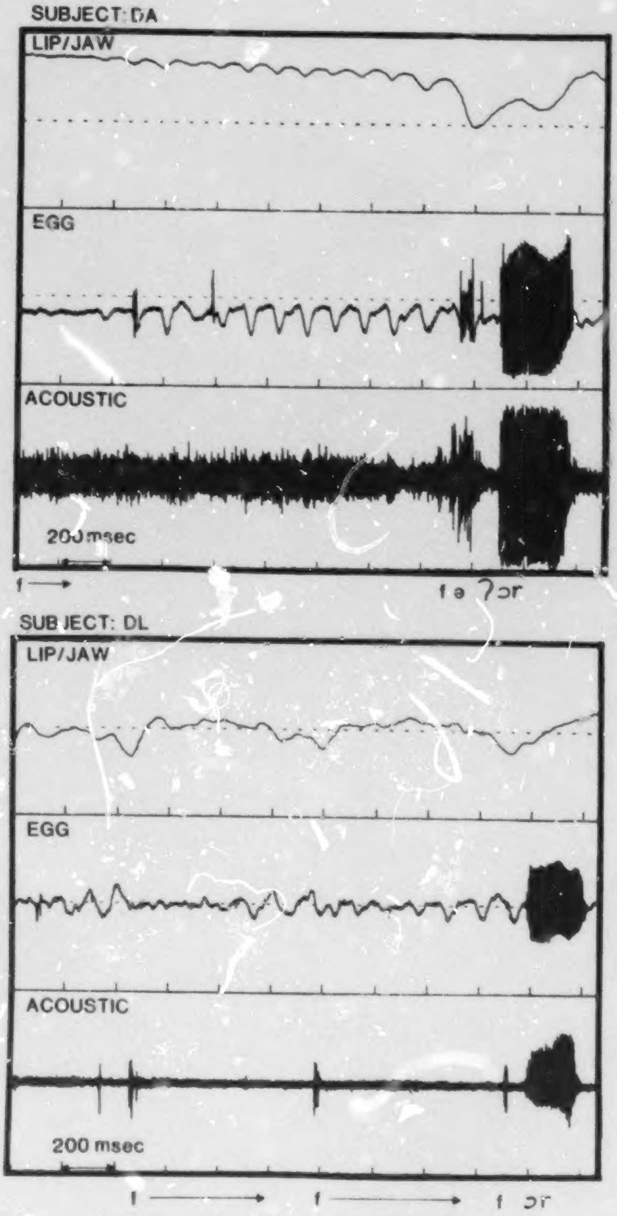


Figure 6. Records of lower lip movement, EGG activity, and the acoustic signals from two subjects. The top part of this figure shows a 9-Hz physiological tremor in both lip and larynx as the subject prolongs [f] in an effort to initiate voicing. The bottom part of the figure shows several repetitions of [f] with lip lowering for each. Superimposed upon these trials is a 7-Hz tremor in both lips and larynx.

posed upon a 1.4-Hz trial frequency, as the subject repeated [f]. Interesting to note here is the normal temporal coordination of the lip/jaw system with the laryngeal adductory system for these repeated trials, even though stuttering is usually considered to be "uncoordinated."

Voice Onset Time

One index of the temporal coordination of laryngeal and supralaryngeal behavior is the measurement of VOT (Lisker & Abramson, 1964) in syllables that consist of a stop and a vowel. Measurements of the time between the burst for /t/ and the onset of the voicing for /u/ in the utterance two were made for all utterances, both stuttered and fluent, in the adaptation task. Figure 7 shows the results. Any utterance that showed aberrant laryngeal activity in the EGG recording was eliminated from the "fluent" category. Thus, the perceptually and physiologically fluent utterances of the mild stutterers were well within normal limits of VOT. Two of the control subjects had considerably longer VOT than the others, with one having a mean VOT of 80 ms and the other 83 ms. They also ranked seventh and eighth, respectively, in syllable rate. There was no significant correlation between VOT scores and rate among the normal speaking group as a whole ($r_s = .43$), but extremely long VOT scores corresponded with the slowest rates. The same finding held for the fluent utterances of the experimental group. The correlation between VOT and rate was low and lacked significance ($r_s = .27$), but at the extremes there was some correspondence in that the subject with the shortest VOT (38 ms) had the fastest speaking rate (when fluent), while the subject with the longest VOT (97 ms) had the slowest speaking rate. The severe stutterers in this study had VOTs that varied depending on whether the block occurred on the utterance two or elsewhere. If the block occurred elsewhere in the series of four digits, VOT on /tu/ fell within normal limits, but if the moment of stuttering fell at the junctive of the voiceless /t/ and the voiced /u/, then VOT was either artificially shortened (as when the subject voiced the stop) or it was extremely long (when voicing became difficult to initiate).

These data do not suggest an overall deficit in VOT among stutterers unless they are stuttering. The Grand Mean for all measures of VOT for control subjects in the utterance /tu/ was 57 ms with a standard deviation of 17 ms, which corresponds closely with the mean VOT of the pooled fluent utterances of stutterers of 56 ms with a standard deviation of 19 ms.

Discussion

When dysfluent, it is in voice initiation that stutterers suffered particular difficulty. Difficulties were manifested in silent blocks, repetitions of the voiceless consonant preceding voicing, or short bursts of voicing that were improperly initiated and were not maintained. After a stuttering block, the most successful strategy for voice initiation was "easy onset of voicing" evidenced by gradual growth of the EGG envelope. After repeated trials, however, the adapted fluent samples were initiated with abrupt EGG envelopes similar to those of the control speakers. VOT measured from the fluent utterances of the stutterers did not significantly differ from that of the controls. Taken together, the results of the VOT analysis and the observations of EGG and acoustic waveforms indicate that, when fluent, stutterers initiate voicing normally.

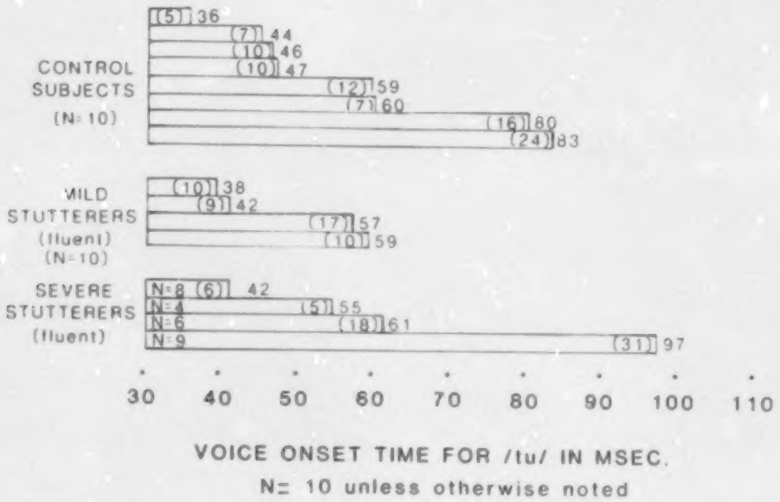


Figure 7. Voice onset time (VOT) as measured from sound spectrograms of the fluent utterances [tu] of the stutterers during the adaptation task and those of the control subjects. Each of the 8 control subjects yielded 10 samples of [tu]. The mean VOT for each subject is noted to the right of each histogram with standard deviations in parentheses. VOT measures for the fluent utterances of stutterers were similar to those for normal speakers. For some of the severe stutterers fewer than 10 fluent samples were obtained since perceptually fluent samples were omitted when abnormal fluctuations in laryngeal impedance preceded voicing.

Although stuttering may reasonably be thought to be a disorder of timing, the obvious temporal irregularities (abnormal VOT, repetitions, and prolongations of sound or silence) may emerge from a problem that has more to do with improper levels of activity than improper timing. The abnormalities of motor coordination seen in stuttering may not be at essence a problem in temporal coordination but rather a problem in the levels of coordinated activity of the many muscles cooperating for a particular function, such as those that set the position and tension of the vocal folds.

Evidence for this theory lies on one hand with the previously noted abnormally high f_0 settings in the aborted voicing trials of some of the stuttering episodes, the less gradual opening phase and less stable open phase of the rapid vocal fold vibrations, all of these factors indicating abnormal stiffness, and on the other hand in the abnormally slow but extreme impedance changes during some of the stuttering episodes indicating wide postural excursions of either too much adduction or too much abduction to permit successful voice initiation. Along with evidence that the settings for the postural and tension prerequisites to voice initiation may be aberrant in stuttering, there is evidence that some temporal coordination is maintained. It is true that VOT as measured in the acoustic signal is abnormal during a stuttering block, but the laryngeal and supralaryngeal systems involved show a remarkable degree of temporal coordination in their movements. The product, the sound, is temporally disorganized due to difficulty in initiating voicing, but the preparatory adjustments are time-locked, and in this sense are well "coordinated." The physiological tremors seen in the laryngeal and lip-jaw records from two of our subjects agree in frequency and tend to be time-locked, and the trials or repetitions demonstrate remarkable temporal bonding of the two systems. It may be that the timing of laryngeal-supralaryngeal coordination is not the parameter at fault in stutterers, rather it may be that levels of the laryngeal activity previous to voice onset or offset are faulty. Zimmermann and Hanley (1983) suggest that in adaptation, background muscle activity in stutterers becomes stabilized as arousal decreases.

When fluent, stutterers yielded VOTs well within normal limits. Reasons that this study found no significant difference while other studies have found longer VOT in fluent utterances of stutterers than in controls may be (1) that the present study used physiological criteria as well as perceptual judgments to categorize an utterance as "fluent" and (2) that repeating utterances until fluent (adaptation) may be a more reliable method of obtaining a fluent sample than picking "fluent" samples out of a corpus of stuttered and fluent speech.

The first report on this experiment (Borden, 1983) suggested that stutterers when they are fluent are similar to their controls in initiating speech. However, in executing a speech task, severe stutterers had a significantly lower speech rate than controls. This finding indicates that severe stutterers may require more time to make the ongoing adjustments and transitions required in speaking fluently. The present study adds support to the first report in that voice initiation seems normal as observed in electroglottograhic waveforms and as VOT measured from spectrograms when stutterers are speaking fluently. When stutterers are dysfluent, however, the folds may not move (the subject with the reversed CV durations), they may go into tremor, or they may exhibit ritualistic patterns involving wide excursions. When voicing is finally initiated successfully after a stuttering block, it is usually by a strategy involving a gradual growth of vibratory amplitude.

These data provide an empirical basis for the use of "easy onset of voicing" techniques in therapy for stutterers. Our observations lead us to caution, however, that easy onset may be revealed more reliably from electroglottographic information than from acoustic waveforms. To the degree that stutterers do initiate voicing normally when fluent, as indicated by the data in this study, another implication for therapy might be that stutterers may profit from enhancing their kinesthetic sense of prevoicing settings when fluent and try to recapture that sense when they are having difficulty in voice initiation.

References

Adams, M. R., & Hayden, P. (1976). The ability of stutterers and nonstutterers to initiate and terminate phonation during production of an isolated vowel. Journal of Speech and Hearing Research, 19, 290-296.

Adams, M. R., & Reis, R. (1974). Influence of the onset of phonation on the frequency of stuttering: A replication and reevaluation. Journal of Speech and Hearing Research, 17, 752-754.

Baer, T., Löfqvist, A., & McGarr, N. S. (1983). Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. Journal of the Acoustical Society of America, 73, 1304-1308.

Bloodstein, O. (1975). *A handbook of stuttering* (Rev. ed.). Chicago: National Easter Seal Society for Crippled Children and Adults.

Borden, G. J. (1983). Initiation versus execution time during manual and oral counting by stutterers. *Journal of Speech and Hearing Research*, *26*, 389-396.

Childers, D. G., Naik, J. M., Larar, J. N., Krishnamurthy, A. K., & Moore, G. P. (1983, May). *Electroglottography, speech, and ultra-high speed cinematography*. Proceedings of the International Conference on Physiology and Biophysics of Voice.

Conture, E. G., McCall, G. N., & Brewer, D. W. (1977). Laryngeal behavior during stuttering. *Journal of Speech and Hearing Research*, *20*, 661-668.

Cross, D. E., & Luper, H. L. (1979). Voice reaction time of stuttering and non-stuttering children and adults. *Journal of Fluency Disorders*, *4*, 59-77.

Desmedt, J. E. (Ed.) (1978). *Physiological tremor, pathological tremors and clonus*. Basel: S. Karger.

Fibiger, S. (1971). Stuttering explained as a physiological tremor. *Quarterly Progress and Status Report* (Speech Transmission Laboratory, Royal Institute of Technology, Department of Speech Communication), 1-23.

Fourcin, A. (1974). Laryngograph examination of vocal fold vibration. In B. Wyke (Ed.), *Ventilatory and phonatory control mechanism* (pp. 315-333). London: Oxford University Press.

Freeman, F. J., & Ushijima, T. (1978). Laryngeal muscle activity during stuttering. *Journal of Speech and Hearing Research*, *21*, 538-562.

Hillman, R. E., & Gilbert, H. (1977). Voice onset time for voiceless stop consonants in the fluent reading of stutterers and nonstutterers. *Journal of the Acoustical Society of America*, *61*, 610-611.

Lippold, O. C. J. (1971). Physiological tremor. *Scientific American*, *224*, 65-73.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384-422.

Metz, D. E., Conture, E. G., & Caruso, A. (1979). Voice onset time, frication, and aspiration during stutterers' fluent speech. *Journal of Speech and Hearing Research*, *22*, 649-656.

Riley, G. (1972). A stuttering severity instrument for children and adults. *Journal of Speech and Hearing Research*, *37*, 314-322.

Rothenberg, M. (1981). Some relations between glottal air flow and vocal fold contact area. In C. L. Ludlow & M. O. Hart (Eds.), *Proceedings of the Conference on the Assessment of Vocal Pathology*, *Asha Reports*, *11*

Ryan, B. (1974). *Programmed therapy for stuttering in children and adults*. Springfield, IL: Charles C. Thomas.

Starkweather, C. W., Hirschman, P., & Tannenbaum, R. S. (1976). Latency of vocalization onset: Stutterers versus nonstutterers. *Journal of Speech and Hearing Research*, *19*, 481-492.

Van Riper, C. (1963). *Speech correction: Principles and methods* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Watson, B. C., & Alfonso, P. J. (1983). Foreperiod and stuttering severity effects on acoustic laryngeal reaction time. *Journal of Fluency Disorders*, *8*, 183-206.

Webster, R. L. (1974). A behavioral analysis of stuttering: Treatment and theory. In K. S. Calhoun, H. E. Adams, & K. M. Mitchell (Eds.), *Innovative treatment methods in psychopathology*. New York: Wiley.

Weiner, A. E. (1978). Vocal therapy for stutterers: A trial program. *Journal of Fluency Disorders*, *3*, 115-126.

Wingate, M. E. (1969). Sound pattern in artificial fluency. Journal of Speech and Hearing Research, 12, 677-686.

Zimmermann, G. N., & Hanley, J. M. (1983). A cinefluorographic investigation of repeated fluent productions of stutterers in an adaptation procedure. Journal of Speech and Hearing Research, 26, 35-42.

Footnote

[1] The main aim of the overall experiment, from which this paper is the second report, was to examine the interaction of respiratory, laryngeal, and supralaryngeal movements of stutterers and their controls during speech. The first report (Borden, 1983) focused on initiation time and execution time for speech and manual counting tasks. The present report focuses on voice onset, and the third report will address coordination.

PHONETIC INFORMATION IS INTEGRATED ACROSS INTERVENING NONLINGUISTIC SOUNDS

D. H. Whalen and Arthur G. Samuel†

Abstract. When the fricative noise of a fricative-vowel syllable is replaced by a noise from a different vocalic context, listeners experience delays in identifying both the fricative and the vowel (Whalen, 1984): Mismatching the information in the fricative noise for vowel and consonant identity with the information in the vocalic segment appears to hamper processing. This effect was argued to be due to phonetic integration of the information relevant to categorization. The present study was intended to eliminate an alternative explanation based on acoustic discontinuities. Noises and vowels were again cross-spliced, but, in addition, the first 60 ms of the vocalic segment (which comprised the consonant-vowel transitions) either had a nonlinguistic noise added to it or was replaced by that noise. The fricative noise and the majority of the vocalic segment were left intact, and both were quite identifiable. Mismatched consonant information caused delays both for original stimuli and for ones with the noise added to the transitions. Mismatched vowel information caused delays for all stimuli, both originals and ones with the noise. Additionally, syllables with a portion replaced by noise took longer to identify than those that had the noise added to them. When asked explicitly to tell the added versions from the replaced, subjects were unable to do so. The results indicate that listeners integrate all relevant information even across a nonlinguistic noise. Replacing the signal completely delayed identifications more than adding the noise to the original signal. This was true despite the fact that the subjects were not aware of any difference.

Phonetic information is spread throughout the acoustic signal. This is true even in the case of fricative-vowel syllables, where it might seem that there are two invariant cues. In such syllables, there are two distinct acoustic segments: a noise that can be identified in isolation as the fricative, and a vocalic segment that can independently specify the vowel. Nonetheless, there is vowel information in the fricative noise (Whalen, 1983; Yeni-Komshian & Soli, 1981), and fricative information in the vocalic formant transitions (Harris, 1958; Mann & Repp, 1980; Whalen, 1981). Thus one of

---

†Department of Psychology, Yale University.

Acknowledgment. This research was supported by NICHD grant HD-01994 to Haskins Laboratories. A portion of this work was presented at the 107th meeting of the Acoustical Society of America, Norfolk, VA, May, 1984. We thank Suzanne Boyce for running the subjects for Experiments 1 and 2. Michael Studdert-Kennedy provided helpful comments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-79/80 (1984)]

the most promising cases of context-independent phonetic cues turns out to vary contextually.

There is also evidence from cross-splicing studies, however, that listeners can detect information specifying the original context of the noise and of the vocalic segment. A series of reaction time studies (Whalen, 1984) indicated that subjects are sensitive to all the information in the syllable. In that work, listeners were presented with edited fricative-vowel stimuli containing mismatches between information in the fricative noise and information in the vocalic segment. Listeners were slower to identify both the consonants and the vowels of the syllables with mismatches, suggesting an attempt to integrate that information, even though the information was not necessary to identify the phones. This was true whether the mismatch was between information about place of articulation in the transitions and in the noise, or between vowel information in the noise and in the vocalic segment itself. It was also true whether the subjects were identifying the vowel or the fricative.

The present experiments were designed to clarify the interpretation of that work. In particular, there was a possibility that some relatively uninteresting psychoacoustic discontinuity in the previous stimuli accounted for the reaction time data. That is, since the stimuli were (digitally) edited, there could have been abrupt changes in the spectrum at the cut, possibly causing a purely auditory disruption of processing. This possibility was less likely in one experiment (Whalen, 1984, Experiment 5) in which, even though the fricative noise was separated from the vocalic segment by 60 ms of silence (thus distancing the two spliced portions), the delay caused by mismatching information remained. However, it is conceivable that the inserted silence failed to displace an auditory trace of the fricative noise. If this trace did not match the vocalic segment, subjects could have perceived a discontinuity. Thus the data do not completely rule out an auditory discontinuity account of the reaction time results.

The present experiments attempt to replicate the slowing effect of mismatches in cases where it is clear that an auditory discontinuity account cannot hold. To that end, the temporal progression of the syllable was left intact (that is, no silence was introduced), but the location of the digital splice coincided with the imposition of a nonlinguistic noise. This noise (either a naturally produced cough or a synthesized buzz) occurred during the vocalic formant transitions, comprising the first 60 ms of the vocalic segment. If the previously obtained delays were mere auditory distractions, then the mismatch effects should disappear--the auditory disturbance at the boundaries of the noise should be the same for syllables with matched and with mismatched fricative noises and vocalic segments. If, however, listeners do in fact integrate information across the whole syllable, then the effect should persist.

Experiment 1

Experiment 1 examined a mismatch of information for fricative place of articulation, between the information in the vocalic formant transitions and that in the noise itself. We will call this a mismatch of consonant information, even though the transitions (as the name implies) provide information about both the consonant and the vowel. The nonlinguistic noise (the natural cough or the synthetic buzz) was introduced in one of two ways. For both matched and mismatched versions, the 60 ms of the vocalic segment which constituted the transitions either had the nonlinguistic noise digitally added (the "added" stimuli), or were replaced by the nonlinguistic noise (the "replaced" stimuli). The added noise was expected to mask the transitions somewhat, presumably reducing the effect of mismatched information if a mere auditory distraction was the cause. However, if the more global, phonetic interpretation is correct, the mismatch should be just as strong when there is noise added to the signal as when the mismatch is the only complicating factor. The replaced stimuli, however, would not have transitions present, and therefore should show no effect of the cross-splicing.

Two different noises were used to reduce the possibility of some unexpected acoustic artifact. We wanted syllables to be perceived as interrupted in a way that allowed what might be called "phonetic" restoration (after Warren's, 1970, phonemic restoration). That is, listeners should be able to assume that there was a signal behind the noise, even in the replaced stimuli. Both noises were primarily aperiodic but with some periodic shaping, a combination most likely to produce phonemic restoration (Samuel, 1981b).

Procedure

Natural tokens of the syllables [sɑ], [ʃɑ], [su], and [ʃu] were recorded by a male speaker of English. (The speaker was not the same as in Whalen, 1984.) The tokens were digitized (20 kHz sampling rate, 9.6 kHz low-pass filtered), and test items were selected so that:

1. All fricative noises were of the same duration (160 ms).
2. All vocalic segments were of the same duration (340 ms).
3. Each syllable token was used either for its fricative noise or for its vocalic segment--thus every test syllable had an electronic splice in it.

Two tokens of each category (e.g., the [s] from [sɑ]) were used.

Two different nonlinguistic noises were used. One was 60 ms of a natually produced cough, while the other was 60 ms of a buzz consisting of a semi-periodic filtering of white noise with peaks at intervals of 500 Hz.

Five copies of each digitized syllable were made. One of these (the "original") was intact except for the digital splice between the fricative noise and the vocalic segment (as described above). Two "added" and two "replaced" versions were constructed: In the "added" versions, the cough noise or buzz noise was added digitally to the first 60 ms of the vocalic segment. In the "replaced" versions, the first 60 ms of the vocalic segment were completly replaced by the cough or buzz.

For all three types of stimuli ("original," "added," and "replaced"), half of the syllables had vocalic segments matched with the fricative noise (e.g., the [u] from [su] paired with an [s] noise) and half had mismatched ones (e.g., the [u] from [ʃu] paired with an [s] noise). Note that when the nonlinguistic noise replaced the first 60 ms of the vocalic segment, there was very little left to be mismatched. That is, even though the rest of the vocalic segment came from an inappropriate syllable, the transitions were, by design, mostly completed by 60 ms. Thus there should not have been much of a phonetic mismatch in the "replaced" stimuli. The first column of Table 1 shows the construction of the matched stimuli, while the second column shows the contruction of the mismatched stimuli. The match/mismatch factor, the five noise conditions (original; added and replaced for two types of noise), and the two tokens of the four fricative and vowel categories result in eighty stimuli.

Table 1

Construction of the Stimuli

Syllable Heard As:	Matched (Exp 1 & 2)	Consonant Mismatch (Exp 1)	Vowel Mismatch (Exp 2)
	noise voc.	noise voc.	noise voc.
"sɑ"	s[ɑ] + [s]ɑ	s[ɑ] + [ʃ]ɑ	s[u] + [s]ɑ
"ʃɑ"	ʃ[ɑ] + [ʃ]ɑ	ʃ[ɑ] + [s]ɑ	ʃ[u] + [ʃ]ɑ
"su"	s[u] + [s]u	s[u] + [ʃ]u	s[ɑ] + [s]u
"ʃu"	ʃ[u] + [ʃ]u	ʃ[u] + [s]u	ʃ[ɑ] + [ʃ]u

Note: Each column presents the syllables used to construct the simulus syllables. The portion of each syllable enclosed in brackets was digitally excised.

In each of two conditions, subjects heard randomized sequences containing five repetitions of each stimulus over headphones. The inter-stimulus interval was 2500 ms. Subjects were asked, in one condition, to identify the vowel ("a" or "u") as quickly as possible. In the other condition, they were asked to identify the consonant ("s" or "sh") as quickly as possible. The order of these conditions was balanced across subjects, as was the determination of which button was pushed by the dominant hand. Responses under 100 ms were counted as mistakes, and the equipment was forced to give up waiting for an answer after 2500 ms. Missing responses and mistakes in identification accounted for 5.0% of the consonant judgments and 3.8% of the vowel judgments. These trials were not included in the reaction time analyses.

The subjects were 20 Yale students who were paid for their participation, all native speakers of English with no reported hearing difficulties.

Results and Discussion

Figure 1 shows the reaction times in Experiment 1 for the first two factors of interest. Overall, mismatches of consonant information, as seen in the left pair of bars, slowed identifications a significant 16 ms, $F(1,19) = 9.97$, $p < .01$. The presence of noise also slowed reaction times, $F(4,76) = 8.19$, $p < .001$, as is seen in the three bars to the right. Adding the noise caused an 8 ms delay, and replacing the noise caused an additional 12 ms delay.

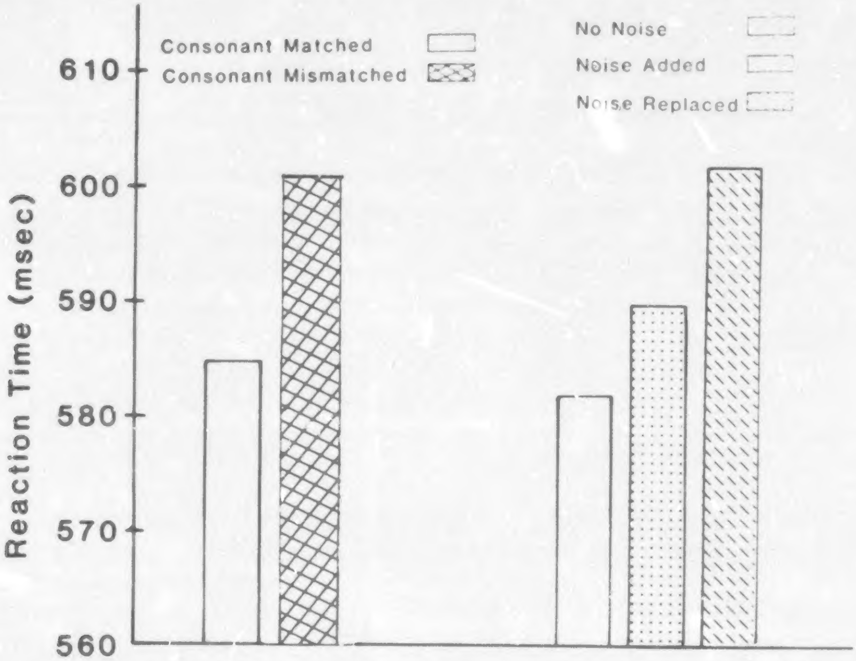


Figure 1. Identification times for stimuli with matched or mismatched consonant information (left pair of bars) and for stimuli with no noise, noise added, or noise replaced (right trio of bars) (Experiment 1).

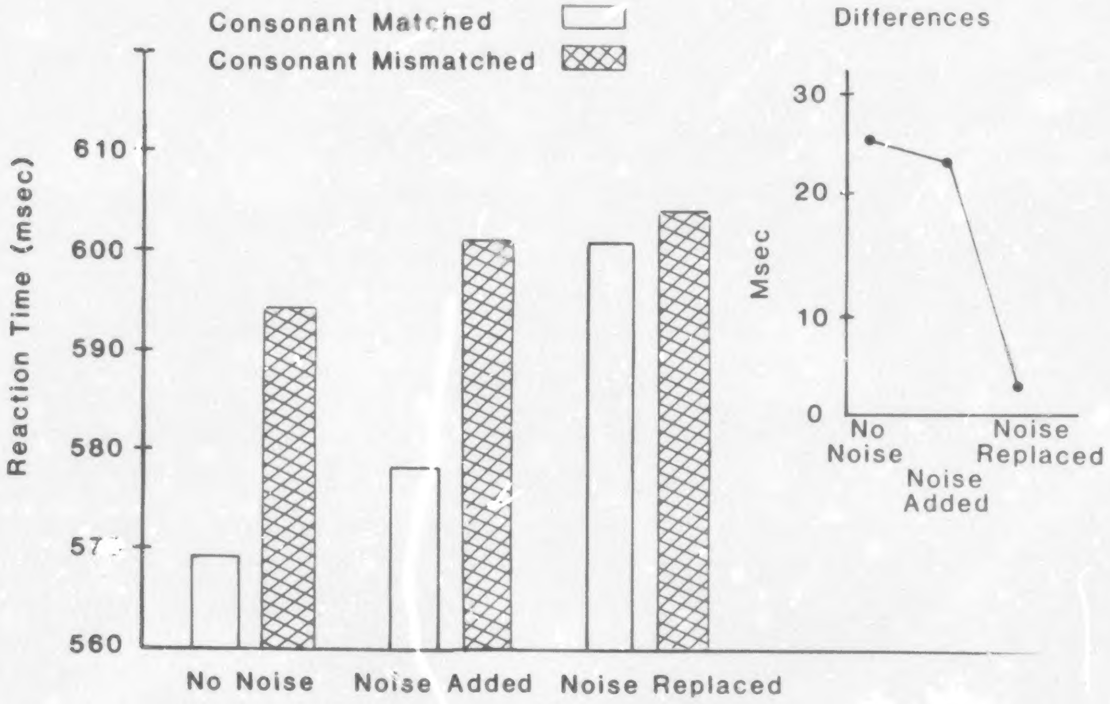


Figure 2. Identification times for stimuli with matched (open bars) or mismatched (cross-hatched bars) consonant information with no noise (left-most pair), noise added (middle pair) or noise replaced (right-most pair) (Experiment 1).

Figure 2 shows the interaction of consonant information mismatch and extraneous noise. In each pair of bars, the open bar shows the mean reaction time to stimuli with matched consonant information. The cross-hatched bar shows the responses to stimuli with mismatched consonant information. The most important result is apparent in the middle pair of bars. Even though both matched and mismatched stimuli include acoustic discontinuities (in the form of the nonlinguistic noises), the mismatch is still robust, *F*(1,19) = 22.32, *p* < .001, for just the "added" stimuli. The comparison of these bars with the two leftmost shows that the addition of the noise slowed judgments an average of 8 ms; the mismatch of transitions added 24 ms whether the noise was present or not.

As can be seen from the rightmost pair of bars, and from the plot of the differences between bars on the right, the difference between matched and mismatched stimuli is negligible in the replaced stimuli (a nonsignificant difference of 2 ms). Not only is there an interaction between added/replaced and match/mismatch, *F*(1,19) = 12.76, *p* < .01, but a separate analysis of the replaced data alone shows no effect of mismatch, F(1,19) = 0.32, n.s. As predicted, there is not enough transitional information left after 60 ms for a mismatch to be detected.

The effect of mismatch was the same whether the consonant or the vowel was identified, *F*(1,19) = 1.97, n.s., for the interaction. Reaction times did not vary due to the type of nonlinguistic noise, *F*(1,19) = 0.61, n.s., nor did type of noise interact with any other factors.

The previously obtained slowing of reaction time with mismatched information was found even when explicitly nonlinguistic (in a sense, purely auditory) discontinuities were present. The effect on identification was not weakened by any masking of the transitions that might have occurred: The phonetic relevance of the transitions was still perceived. It is still conceivable that there are two auditory discontinuities at work (the transitions and the nonlinguistic noises) and that they do not interfere with each other. Experiment 2 examines a situation where this interpretation is not possible.

Note that the identification times for the replaced stimuli are essentially the same as for the mismatched added stimuli (see Figure 2): The delay caused by a mismatch is the same as the delay caused by the absence of the original signal. One interpretation of this is that appropriate transitions facilitate identification, and that mismatched transitions are no worse than having no transitions at all. Alternatively, the similarity in mean reaction times might be coincidental. Experiment 2 provides an opportunity to test these alternatives while examining the effect of mismatching vowel information.

Experiment 2

Experiment 2 mismatched the vowel information in the fricative noises with that of the vocalic segment. Manipulations similar to those of Experiment 1 were carried out, but with a different expectation: Mismatches of phonetic information should show up even in the replaced stimuli. This is based on the fact that the vowel mismatch does not depend just on the first 60 ms of the vocalic segment, but is instead present throughout the noise, on the one hand, and the vocalic segment, on the other.

Procedure

The syllable pieces of Experiment 1 were again used in Experiment 2, although the combinations for the mismatched stimuli were different. The matched stimuli were identical (see Column 1 in Table 1). The mismatched syllables are outlined in the third column of Table 1. The transitions were always appropriate to the fricative, i.e., the consonant information was matched. The same five noise conditions as in Experiment 1 were used in Experiment 2: no noise, cough or buzz added to the first 60 ms of the vocalic segment, or cough or buzz replacing those 60 ms.

The stimuli were presented as before, with the two conditions of consonant identification and vowel identification, each presented as a separate block. Missing responses and mistakes in identification accounted for 4.7% of the consonant judgments and 3.1% of the vowel judgments. These trials were excluded from further analysis.

The subjects were 20 Yale students who were paid for their participation. All were native speakers of English with no reported hearing difficulties. Half had participated in Experiment 1.

Results and Discussion

Figure 3 presents the results of mismatching vowel information and for including noise in the stimuli. The two bars at the left indicate that mismatching vowel information had a significant slowing effect of 24 ms, $F(1,19) = 46.90$, $p < .001$. The three bars on the right indicate that adding noise slowed judgments by 29 ms, while replacing part of the syllable with noise slowed judgments by an additional 15 ms, $F(4,76) = 29.73$, $p < .001$. All three of these categories were significantly different from each other.

Figure 4 shows the results by both match and noise condition. In each pair of bars, the open bar shows the mean reaction time to stimuli with matched vowel information. The cross-hatched bar shows the responses to stimuli with mismatched vowel information. Unlike Experiment 1, vowel mismatches caused delays in each case; the effect of mismatches did not differ across these conditions, $F(4,76) = 0.27$, n.s. If anything, these delays increased with the presence of noise, as is shown by the plot on the right. This plot shows the differences between the matched and mismatched stimuli for the no noise, noise added and noise replaced stimuli respectively from left to right.

There was one interaction between the match/mismatch factor and the category identified (consonant or vowel). The mismatch effect was approximately twice as large when the vowel was identified (15 ms for consonant identification, 31 for vowel, $F(1,19) = 6.78$, $p < .05$. A separate analysis of the consonant identification data alone shows that the effect of mismatch was still significant, $F(1,19) = 9.57$, $p < .01$.

The main effect of noise type was not significant, $F(1,19) = 1.12$, n.s., nor did it enter into any significant interactions.

As in Whalen (1984), mismatching the rather weak vowel information in the fricative noise with the more powerful information in the vocalic segment slowed phonetic judgments. Even though a nonlinguistic noise indicated that the signal had been corrupted, listeners were still affected by mismatches be-

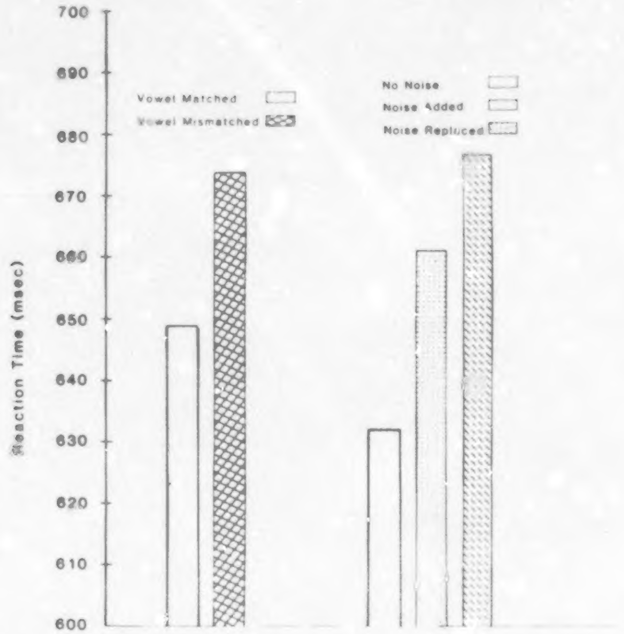


Figure 3. Identification times for stimuli with matched or mismatched vowel information (left pair of bars) and for stimuli with no noise, noise added, or noise replaced (right trio of bars) (Experiment 2).

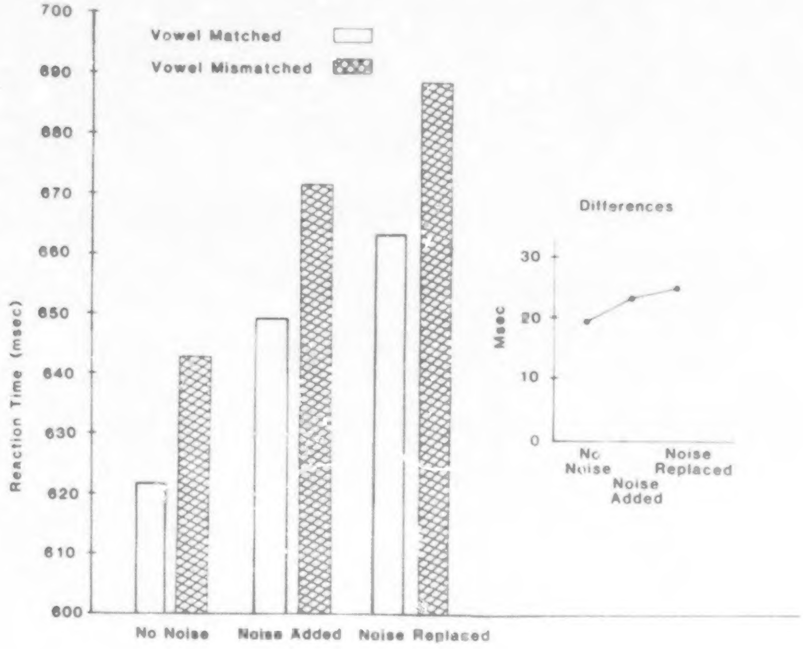


Figure 4. Identification times for stimuli with matched (open bars) or mismatched (cross-hatched bars) vowel information with no noise (left-most pair), noise added (middle pair) or noise replaced (right-most pair) (Experiment 2).

tween two temporally separated portions of the utterance. The present experiment is particularly interesting because the information critical to the mismatch was not removed when the nonlinguistic noise replaced the transitions (as it was in Experiment 1). The "replaced" stimuli in Experiment 2 demonstrated that even when *all* tokens include significant acoustic discontinuities, the disruption due to mismatching phonetic information persists: The identification delays are due to an impairment of the process that integrates relevant information, not to any simple distractions caused by auditory discontinuities.

One difference between the two experiments is the absolute amount of time it took for the phonetic decisions. Subjects were, on the whole, 68 ms slower in Experiment 2 than in Experiment 1. An analysis (with the factors used before plus the factor of Experiment) of the ten subjects who participated in both experiments shows that the difference is a real one; the effect of Experiment was reliable, $F(1,9) = 8.91$, $p < .05$. The only interaction of the Experiment factor was with Mismatch and Noise, which was expected, since the effect of mismatches disappeared for the replaced versions in Experiment 1 but not in Experiment 2. In the first experiment, 30% of the stimuli had detectable mismatches of phonetic information, while in the second, 50% did. This increase of conflicting information probably resulted in more cautious identifications, slowing down responses.

The fact that the two delaying effects, mismatches of vowel information and the addition of the two types of noise, were independent allows us to choose between two explanations proposed for the results of Experiment 1. In that experiment, it seemed either that mismatched transitions slowed identification, or that the availability of appropriate information speeded identification. The similarity of identification times for syllables with mismatched information to those where the noise replaced the transitions left both possibilities open. As can be seen in Figure 4, mismatched information slowed identifications whether nonlinguistic noise was present or not. These results indicate that both the mismatches and the nonlinguistic noise have an interfering effect on identification times.

Experiment 3

The first two experiments have shown that subjects are sensitive to whether the signal is intact or not: In both, replaced stimuli produced significantly slower reaction times than added stimuli. One possible explanation for this effect is that the replaced items are heard as interrupted or discontinuous and that this distracts the subjects enough to slow them down. A more likely explanation, given that phonetic integration occurs across the noise, is that the perceptual system expects to find the signal even when nonlinguistic noises are present, and that perceptual processing is slowed when this expectation is not met. Experiment 3 tests whether there are noticeable differences between added and replaced stimuli that would support the "distracting" hypothesis. The test involves explicitly asking the subjects to discriminate between added and replaced stimuli. If the subjects are being distracted by the replacement of the signal, then added and replaced stimuli should be discriminable.

Procedure

The stimuli were the "added" and "replaced" items used in the first two experiments. Ninety-six tokens were used in Experiment 3, representing the crossing of four factors: (1) buzz versus cough as extraneous noise, (2) added versus replaced, (3) matched, consonant mismatched, or vowel mismatched, and (4) tokens. The last factor, tokens, represents the eight examples within each cell of the design, and includes two instances each of /sɑ/, /ʃɑ/, /su/, and /ʃu/.

The stimuli used in Experiments 1 and 2 were recorded on audiotape and digitized on another computer system, using high-quality audio components and a 12-bit A/D converter. The sampling rate was 20 kHz, with 9.6 kHz low-pass filtering.

The entire stimulus set of 96 items was presented twice. The first 48 stimuli spanned all of the factors just described except "added versus replaced." The form of each token ("added" versus "replaced") was randomly selected. The second set of 48 stimuli included the "other" form ("replaced" if the "added" form of a token had already been presented, and vice versa). The second pass through the 96 stimuli used the same procedure. Each group of 48 tokens was randomly ordered.

Subjects were told that they would be hearing "sa," "sha," "su," and "shu," with some noise present during each syllable. It was explained that the noise would occur "where the consonant met the vowel," and that the noise would either replace a small bit of the syllable, or be superimposed on it. Subjects were instructed to press one button on a computer terminal if they thought the noise replaced part of a syllable, and another button if they thought the noise was superimposed.

The presentation of stimuli was subject-paced: Approximately one second after a subject's response was received, the next stimulus was presented. The entire procedure took approximately 15 minutes.

Twelve individuals served as subjects in Experiment 3. They were recruited through sign-up sheets posted at Yale University, and were paid for their participation. All were native English speakers with no reported hearing problems. Half had previously participated in another study in which they made similar judgments.

Results and Discussion

The central question of Experiment 3 is whether listeners can discriminate the "added" and "replaced" versions of the syllables. To answer this question, the percentage of correct responses was calculated for each subject, broken down by matching condition (match, consonant mismatch, vowel mismatch), extraneous noise (buzz or cough), and stimulus form ("added" or "replaced"). These percentages were used to calculate the signal detection parameter d'. This bias-free measure of discrimination performance was computed for each of the six cells defined by the crossing of the three matching conditions and the two extraneous noises. These values were submitted to a two-factor analysis of variance (matching condition X extraneous noise).

The results of this analysis can be summarized very simply: Subjects are utterly unable to discriminate "added" and "replaced" stimuli. In signal detection analyses, a d' of 0 indicates no discriminability, with increasing values reflecting an ability to discriminate. The obtained grand mean d' was -0.003, indicating that the "added" and "replaced" stimuli could not be discriminated at all. Given this, it should not be surprising that neither extraneous noise type, $\underline{F}(1,11) < 1$, nor matching condition, $\underline{F}(2,22) = 2.84$, n.s., made a significant difference; their interaction was similarly inconsequential, $\underline{F}(2,22) = 1.31$, n.s.

What makes this null result of interest is that the "added" and "replaced" stimuli produced significantly different reaction times in Experiments 1 and 2. We thus have a situation in which a manipulation that is totally unavailable to consciousness produces reliable differences in processing time. The extra acoustic discontinuity produced by the replacement manipulation is sufficient to slow down identification of the speech signal (Experiments 1 and 2), but is not discriminable from the mere addition of noise (Experiment 3).

The inability of listeners to discriminate between the "added" and "replaced" items when they are explicitly asked to do so is reminiscent of results obtained in studies of the phonemic restoration effect (Samuel, 1981a). An important difference to note, however, is that in studies of restoration, care is taken to remove all local cues to a phone; if the stretch of speech immediately before or immediately after the replacement locus is played, the relevant phone will not be heard. In the present study, both the fricative and the vowel are perfectly intelligible in isolation; only the transitions are replaced (or have noise added). Thus, there is not enough evidence to tell whether the present results reflect some sort of restoration. A better analogy might be to the classic categorical perception findings (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In these studies listeners also fail to discriminate between acoustically different tokens (ones within a phonemic category). Moreover, just as in the present study, reaction time analyses of identification times reveal differences between these indiscriminable items (Pisoni & Tash, 1974). The reaction time analyses thus provide insights into the processing of speech that cannot be revealed in overt discrimination tasks.

General Discussion

The phonetic mismatch effects of Whalen (1984) were successfully replicated, even with stimuli containing a nonlinguistic noise, inviting the auditory system to block integratation of portions of the signal. The present study also shows that having the original signal behind the noise is less disruptive than replacing the signal altogether. This indicates that the perceptual system looks for coherence even within competing noise. The results of this search for coherence are not available to consciousness, as is shown in Experiment 3.

It appears then that listeners are indeed sensitive to all phonetic information given them, and that the delays caused by mismatches, even those that cannot be readily heard, are due to increased phonetic processing. Even when the subject is given every excuse for failing to integrate, as when a nonlinguistic noise occurs in the middle of the signal, she still does integrate. The mismatch adds just as much time to the perceptual process whether the extraneous noise is present or not. This indicates that the previously obtained result is not simply a short-term psycho-acoustic disruption but is sustained over a relatively long stretch. Whether the information stored is acoustic or (weakly) categorical remains to be seen.

References

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, *1*, 1-7.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431-461.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*, 213-228.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*, 285-290.

Samuel, A. G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*, 474-494.

Samuel, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1124-1131.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*, 392-393.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [š]-[s] boundary. *Journal of the Acoustical Society of America*, *69*, 275-282.

Whalen, D. H. (1983). Vowel information in postvocalic fricative noises. *Language and Speech*, *26*, 91-100.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, *35*, 49-64.

Yeni-Komshian, G. H., & Soli, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, *70*, 966-975.

PARAMETERS OF SPECTRAL/TEMPORAL FUSION IN SPEECH PERCEPTION*

Bruno H. Repp and Shlomo Bentin†

Abstract. When the distinctive formant transition of a synthetic syllable is presented to one ear while the remainder (the "base") is presented to the opposite ear, listeners report hearing the original syllable in the ear receiving the base--a phenomenon called "spectral/temporal fusion" by Cutting (1976). We have found that the mere onset (i.e., the first pitch pulse, 10 ms in duration) of an isolated, contralateral third-formant (F3) transition can be sufficient to cue the /da/-/ga/ distinction in this way. We also varied the relative onset times of isolated F3 and base, and compared three types of F3 segments (50-ms time-varying, 50-ms constant, 10-ms onset) under both dichotic and diotic presentation. Time-varying F3 segments were superior to constant ones, especially when they lagged behind the base. Diotic performance exceeded dichotic performance, but only when F3 preceded the base, suggesting that upward spread of masking occurred in diotic presentation when F3 coincided with energy in the lower formants. Perhaps most interestingly, subjects' tolerance of temporal asynchrony (roughly ±50 ms) was about the same in dichotic and diotic conditions, suggesting that the temporal integration mechanism that combines phonetic information from the isolated F3 segment and the base operates similarly in both conditions.

It has long been known that perceptual fusion results when the first formant (F1) of a synthetic speech signal is presented to one ear while the higher formants are simultaneously presented to the other ear (Broadbent, 1955; Broadbent & Ladefoged, 1957). In this situation, listeners perceive a single fused stimulus localized toward the side of F1 (cf. Darwin, Howell, & Brady, 1978). A variant of this paradigm was introduced by Rand (1974), who presented only the time-varying F2 and F3 transitions of CV syllables to one ear while F1 and the steady-state portions of F2 and F3 were presented to the opposite ear. The perceptual fusion that occurs in this situation has been labeled "spectral/temporal fusion" by Cutting (1976).

Spectral/temporal fusion has received considerable attention in recent years. Research on "duplex perception" (Bentin & Mann, 1983; Liberman, 1979;

---

*Perception & Psychophysics, in press.

†Now at Department of Neurology, Aranne Laboratory of Human Psychophysiology, Hadassah Hospital, Jerusalem, Israel.

Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. Shlomo Bentin was supported by a stipend from the Jesselson foundation. We are grateful to Alvin Liberman, Michael Studdert-Kennedy, and two reviewers for their helpful comments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-79/80 (1984)]

Liberman, Isenberg, & Rakerd, 1981; Mann & Liberman, 1983; Nusbaum, Schwab, & Sawusch, 1983; Repp, Milburn, & Ashkenas, 1983) has focused on the fact that, simultaneously with the speech, the isolated formant transition is perceived as a nonspeech "chirp." Thus the isolated transition contributes to phonetic and nonphonetic percepts at the same time, which has been interpreted as evidence for the simultaneous operation of a speech-specific and a general auditory mode of perception (Liberman, 1982; Liberman et al., 1981; Mann & Liberman, 1983). Recent studies have shown that the speech and nonspeech percepts in this situation are affected in different degrees by manipulations such as masking or attenuation of the distinctive isolated transition (Bentin & Mann, 1983).

In the present studies, we are not directly concerned with duplex perception as such. Rather, we focus on the speech percept only and examine some of the factors that may limit the occurrence of fusion in this special situation. By "fusion" we mean here the contribution of the isolated transition to speech identification. The strict definition of fusion as a single stimulus percept from two separate inputs clearly does not apply in duplex perception. In Experiment 1, we examine how long the distinctive isolated formant transition must be to enable listeners to discriminate between two alternative syllables when attending to the ear receiving the nondistinctive base. Experiment 2 is a parametric study of the effects of temporal asynchrony on spectral/temporal fusion, including comparisons of dynamic and static "transitions," and of dichotic versus diotic presentation.

Experiment 1

All previous studies of spectral/temporal fusion have followed the standard paradigm described above. In each case, a complete formant transition was presented to the ear contralateral to the base, although the duration of the isolated transition varied from 30 to 70 ms across different studies. In the present study, we wished to determine, first, whether the full transition is needed to make the speech distinction, or whether a truncated version or even just the onset of the transition would suffice. Second, we asked whether the presence of the steady-state continuation of the same formant in the base is a necessary condition for spectral/temporal fusion to occur. The second half of the term, "spectral/temporal," suggests that an affirmative answer was assumed by Cutting (1976). To test this inference, we omitted from the base the steady-state resonance following the critical transition, expecting (on the basis of pilot observations) that fusion would nevertheless be obtained. (A direct comparison of conditions with and without this steady-state formant in the base was conducted in Experiment 2.)

The materials used were the syllables /da/ and /ga/, synthesized so as to differ only in the F3 transition. Earlier studies have obtained strong spectral/temporal fusion with similar stimuli (Mann & Liberman, 1983; Repp et al., 1983). The experimental manipulation in Experiment 1, then, was to reduce the duration of the isolated F3 transition (appropriate for either /da/ or /ga/) until only its onset (i.e., the first pitch pulse) remained, while a constant two-formant base was presented in synchrony to the opposite ear. Spectral/temporal fusion was assessed in terms of subjects' ability to distinguish /da/ and /ga/ in the ear receiving the base.

Methods

Subjects. Twelve subjects (three males, nine females) were tested. They were all Yale undergraduates and were paid for their participation.

Stimuli. The stimuli were three-formant synthetic approximations of the syllables /da/ and /ga/, produced on the parallel software synthesizer at Haskins Laboratories, as illustrated schematically in Figure 1. The first two formants were identical in both syllables, and constituted the "base." The duration of the base was 250 ms with a 50 ms amplitude ramp at onset and a constant fundamental frequency of 100 Hz for the first 100 ms, followed by a linear decrease to 80 Hz at offset. The first formant began at 279 Hz and increased linearly in frequency during the first 50 ms to a steady state of 765 Hz. The second formant began at 1650 Hz and decreased linearly in frequency during the first 50 ms to a steady state of 1230 Hz. The base by itself is perceived as either /da/ or /ga/ or as ambiguous, depending on the listener. The /da/ third-formant transition, originally 50 ms (5 pitch pulses) in duration, began nominally at 2800 Hz and decreased linearly in frequency to 2550 Hz, while the /ga/ transition began nominally at 1800 Hz and increased linearly in frequency to 2550 Hz. (These are the "dynamic" transitions in Figure 1; the actual F3 frequencies in the first pitch pulse were 2775 and 1875 Hz, respectively--see caption to Figure 1.) Five transition durations were used, as indicated by the tick marks in Figure 1: 50, 40, 30, 20, and 10 ms (5, 4, 3, 2, and 1 pitch pulses, respectively). Since the frequency trajectory was not changed, the shorter transitions had offset frequencies increasingly closer to the onset frequencies.

The stimuli were recorded onto magnetic tape, with the isolated F3 transitions on one channel and the onset-aligned, constant base on the other. There were 240 stimuli altogether: 24 repetitions of the /da/ and /ga/ transitions at each of 5 durations. The stimuli were arranged in 5 randomized sequences, with ISIs of 2.5 s between stimuli and longer intervals between sequences.

Procedure. The tapes were presented at a comfortable intensity over TDH-39 earphones in a quiet room. The base was always in the left ear and the F3 transition was in the right ear. (No pronounced ear asymmetries have been observed in this task.) Subjects were instructed to listen to their left ear and to identify the syllables in writing as beginning with either "d" or "g."

Results and Discussion

Performance for 50-, 40-, and 30-ms transitions was nearly perfect: 96, 97, and 98 percent correct, respectively. For 20-ms transitions performance dropped to 91 percent correct, and for 10-ms transition onsets, to 84 percent correct. Individual subjects' scores in the last condition ranged from 66 to 96 percent correct. Thus, although there was some loss in accuracy, even the 10-ms single pitch-pulse transition onsets were sufficient to distinguish /da/ and /ga/ in the opposite ear. Accordingly, time-varying frequency information in F3 does not seem essential either for this particular phonetic distinction or for spectral/temporal fusion to occur.

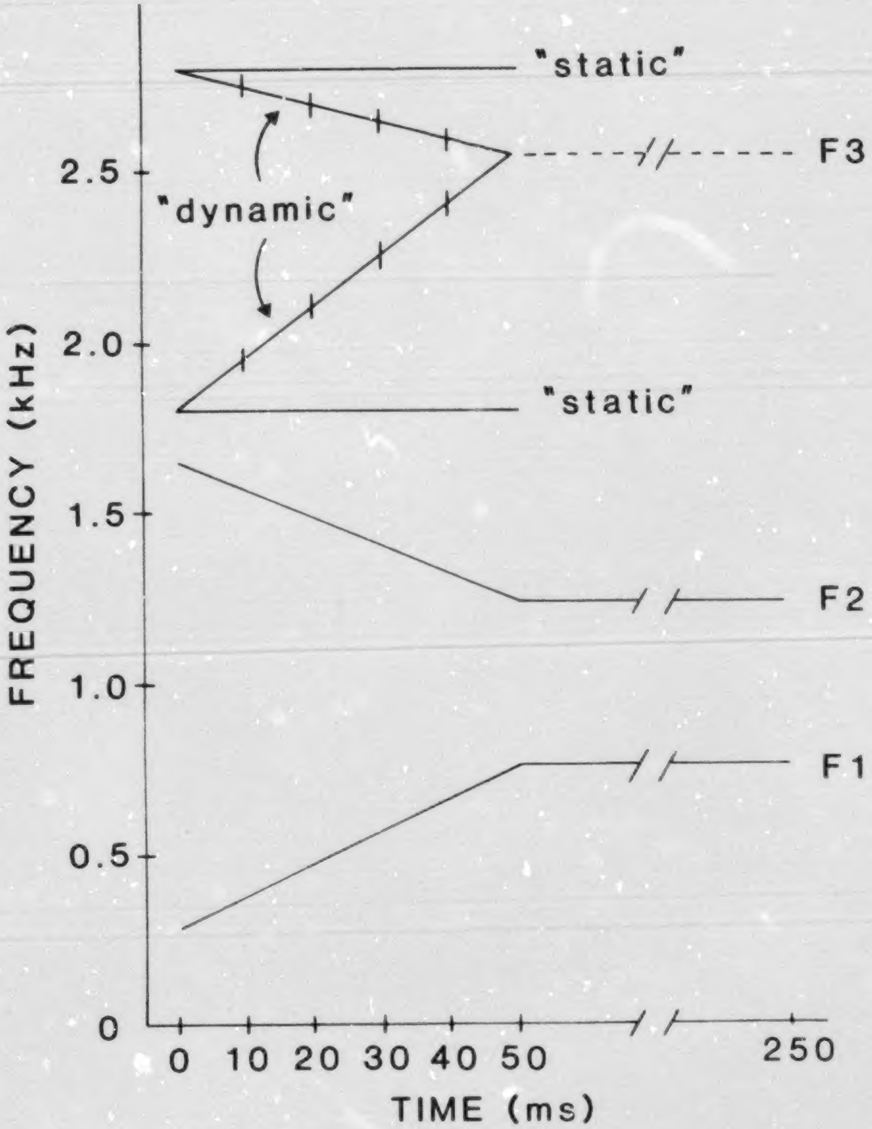


Figure 1. Schematic illustration of the center frequencies of the first three formants in the stimuli of Experiments 1 and 2. All formant transitions are drawn as idealized linear functions connecting the nominal frequencies used in synthesis. The formant frequencies were actually constant within each pitch pulse at values halfway between the nominal onset and offset frequencies for that 10-ms period. The "dynamic" transitions were used in both experiments; the tick marks indicate the shortening manipulation in Experiment 1. The "static" F3 segments were used in Experiment 2 only. The dashed line represents the F3 steady state present in the base on half of the trials in Experiment 2.

In addition, it is clear that the absence of the F3 steady state in the base did not prevent fusion. Since temporal continuity in the relevant frequency band thus seems to contribute little (see also Experiment 2), spectral/temporal fusion appears to be just a special case of spectral fusion (Cutting's, 1976, term for the fusion of complete formants presented simultaneously to different ears). The difference lies in that only the former situation gives rise to a duplex percept (syllable and "chirp"); the mechanism that reconstitutes the speech percept from separate components, however, seems to be the same.

It might be argued that the subjects accomplished their task by paying attention to the chirp-like isolated transition and responding "g" when the chirp was low-pitched and "d" when it was high-pitched (cf. Nusbaum et al., 1983). Even though no catch trials were employed in the present study, this possibility is virtually ruled out by previous evidence that (1) subjects do attend to the ear receiving the base when instructed to do so (Mann & Liberman, 1983; Repp et al., 1983), and (2) they are unable to associate isolated F3 chirps consistently with the response categories "d" and "g" (Repp et al., 1983). Moreover, all listeners agree that the syllables in the ear receiving the base really do sound alternately like /da/ or /ga/. Therefore, the present subjects' responses almost certainly reflect the combination of information from the two ears.

It may be noted that a 10-ms F3 onset is not only devoid of time-varying information but is also nonperiodic, consisting only of a single glottal cycle. By itself, it sounds like a click. Informally, we have confirmed that fusion is also obtained when this 10-ms pitch pulse is replaced with a 10-ms burst of noise with the same spectral envelope, generated by the aperiodic source of the synthesizer. This observation reveals a possible similarity with a phenomenon reported by Pastore, Szczesiul, Rosenblum, and Schmuckler (1982), who found that a burst of filtered white noise changed the perception of a contralateral /pa/ to /ta/. These findings indicate that dichotic integration of phonetic information can occur even if the signal in one ear is periodic and the other is not. It is not clear whether such phenomena should be attributed to general processes of auditory fusion. Rather, they may constitute evidence for a central phonetic decision mechanism that operates on inputs from both ears.

Experiment 2

To explore in more detail the parameters of spectral/temporal fusion, we conducted a multifactorial experiment including four independent variables: (1) A range of onset asynchronies between the isolated F3 segment and the base; (2) dichotic versus diotic presentation; (3) static (constant frequency) versus dynamic (time-varying frequency) F3 segments, and (4) bases with and without a steady-state F3.

Effects of stimulus onset asynchrony (SOA) on spectral/temporal fusion were studied by Cutting (1976) with synthetic two-formant stimuli. The isolated F2 transition was 70 ms in duration. Cutting used transition-base lead and lag times of up to 160 ms, spaced in logarithmic steps, but reported his results averaged over leads and lags, since he found no significant asymmetry. As expected, speech identification performance dropped as SOA increased. However, performance was still slightly above chance even at the longest interval (160 ms), although the statistical significance of this finding was not determined. The longest interval at which performance was substantially above chance was 40 ms.

In a recent study, Bentin and Mann (1983; Exp. 1) used SOAs of up to 100 ms with two-formant syllables similar to Cutting's, although the transitions were only 50 ms in duration. Only lead times were used; that is, the F3 segment always preceded the base. Subjects' performance declined steadily with increasing SOA, but was still above chance at the 100-ms interval. These results are consistent with Cutting's in that they suggest a considerable tolerance of temporal asynchrony in spectral/temporal fusion.

In the present study we sought to replicate these findings with stimuli distinguished by a difference in the F3 transition. Particular attention was given to possible performance asymmetries between lead and lag times. Cutting's (1976) negative finding notwithstanding, such asymmetries might be predicted on at least two grounds. First, when the F3 segment lags behind the onset of the base and thus coincides with the vowel, it may suffer some contralateral simultaneous masking that is absent when the F3 segment precedes the base. Second, when the F3 segment lags behind, listeners may conceivably be able to classify the base phonetically before processing the F3 segment. Both considerations predict stronger fusion when the F3 segment leads the base than when it lags behind. On the other hand, one might also predict the opposite: It is known that, in auditory perception, the terminal frequency of a tone glide is more salient than its initial frequency (Nábělek, Nábělek, & Hirsh, 1970; Schwab, 1981). If a leading F3 segment is retained in auditory memory before it is integrated with the base, its distinctiveness might be reduced because full /da/ and /ga/ transitions have the same terminal frequency. This may confer a relative advantage on lagging F3 segments, which need not be stored in auditory memory.

A second comparison in Experiment 2 concerned dichotic versus diotic presentation of the stimulus components. Rand (1974) conducted such a comparison for onset-synchronous transition and base and found better speech discrimination in the dichotic condition. He attributed this to simultaneous masking of higher by lower formants in the diotic condition, and to release from this form of peripheral upward spread of masking in the dichotic condition. Subsequent studies (e.g., Danaher & Pickett, 1975; Nye, Nearey, & Rand, 1974; Nearey & Levitt, 1974) have replicated this difference, although there are also negative findings in the literature (Nusbaum et al., 1983; Repp et al., 1983). This is the first study to vary SOA in such a comparison. If upward spread of masking operates, then the advantage of dichotic over diotic performance should hold at all lag times, as long as the F3 segment coincides with the base. However, no such difference should exist at lead times, unless there is significant peripheral backward masking of the F3 segment by the base, which seems unlikely.

Another question of interest was whether listeners would be equally tolerant of stimulus onset asynchronies in diotic and in dichotic presentation. Presented monotically or diotically, onset-synchronous diotic transition and base constitute, of course, an intact syllable. It has not been attempted previously to advance or delay the isolated transition with respect to the base when both occur in the same channel. At least one dichotic fusion phenomenon (the influence of a contralateral white noise burst on the perceived place of articulation of a stop consonant) does not seem to occur when the stimulus components are presented diotically (Pastore et al., 1982). We considered it possible that fusion of transition and base in the diotic condition might be restricted to short SOAs, where there is physical overlap, whereas in the dichotic condition subjects might be less sensitive to temporal asynchronies.

A third comparison of interest concerned the nature of the F3 segment conveying the distinctive information. Three kinds of F3 segments were compared: (1) standard 50-ms time-varying ("dynamic") F3 transitions; (2) short 10-ms onsets (as in Experiment 1); and (3) 50-ms constant ("static") F3 segments, which were obtained by extending the transition onset frequencies, as illustrated in Figure 1. The static F3 segments were of special interest: First, would they be sufficient to cue the /da/-/ga/ distinction? (The effectiveness of the short F3 segments in Experiment 1 suggests a positive answer.) Second, would they be as effective as dynamic F3 segments, or does the dynamic information convey additional phonetic distinctiveness? Third, the static F3 segments for /da/ and /ga/ have distinctive terminal (as well as initial) frequencies, which may be an advantage at F3 lead times. Up to a lead time of 40 ms, the distinctive end of a static F3 segment actually still overlaps with the onset of the base. As a result, performance at short lead times may be better for static than for dynamic F3 segments, unless the distinctive phonetic information derives strictly from F3 onset and physical overlap is irrelevant. Comparisons with the short F3 segment should also be enlightening in that regard, although the short duration of this stimulus entails a loss in energy and a consequent decrement in discriminability.

In addition to these three major factors (SOA, mode of presentation, and type of F3 segment), the experiment also included a comparison of bases with and without an F3 steady state. Since Experiment 1 had shown strong fusion in the absence of an F3 steady state, little effect of this last factor was expected.

Methods

Subjects. Twelve paid volunteers participated, six men and six women. Five of them had been subjects in Experiment 1. Of the other seven, two had to be replaced because of exceedingly poor performance.

Stimuli. The basic stimuli were the same as in Experiment 1. In addition to the base used there, a second base was used that included a steady-state F3 at 2550 Hz, starting 50 ms after the onset of F1 and F2, at the same time as the steady states of these formants. (The vowel had very nearly the same quality with and without F3.) There were three kinds of F3 segments: The dynamic (50 ms) and short (10 ms) versions corresponded to the extremes of transition duration used in Experiment 1; the static (50 ms) F3 segments were synthesized at constant frequencies corresponding to the nominal onset frequencies of the dynamic segments (see Figure 1).

Three stimulus tapes were recorded, each corresponding to a different type of F3 segment. Each tape contained 10 blocks of 22 stimuli, each block being a randomization of the two F3 segments for /da/ and /ga/ recorded on one track, at 11 different SOAs in relation to the base on the other track. The 11 SOAs were: -100, -70, -40, -20, -10, 0, 10, 20, 40, 70, and 100 ms; a negative SOA means that the F3 segment led the base. In addition, odd-numbered blocks contained the base without F3, while even-numbered blocks contained the base with a steady-state F3. The ISI was 2 s, and there were 6 s between blocks.

Design and procedure. Each of the three stimulus tapes was presented in two conditions: dichotic and diotic. All six conditions were presented in a single session. The order of conditions was strictly counterbalanced across subjects, with the constraint that all diotic conditions either preceded or followed all dichotic conditions.

A brief familiarization sequence with dynamic F3 segments at SOA=0 was presented at the beginning of the session. This sequence included 10 stimuli in which /da/ and /ga/ alternated, followed by a random arrangement of 20 stimuli. The sequence was first presenteally and then dichotically. The subjects tried to identify the syllables and were given feedback after the sequence. If more than a few errors were committed, the sequence was presented a second time.

Subjects were run individually under the same conditions as in Experiment 1. The tape recorder channels were calibrated for equal intensity of a repeated vowel. Diotic presentation was achieved by mixing the two channels together and feeding the result to both earphone channels. No intensity adjustment was made; because of the relative weakness of the F3 segment, the increase in the total amplitude of the mixed syllables over the isolated base was minimal. In the dichotic conditions, the F3 segment was presented to the right ear for half of the subjects and to the left ear for the other half.

The structure of the stimuli and of the test tapes was explained to the subjects in advance. They were asked not to rely on the high or low pitch of the F3 segment and to focus their attention on the speech percept only. A forced choice between "d" and "g" responses was required for each stimulus.

Results

The main results are shown in Figure 2, where the percentage of correct consonant identifications is plotted as a function of SOA (abscissa), type of F3 segment (separate functions), and presentation condition (separate panels). A 5-way repeated-measures analysis of variance was conducted that included, in addition to the three factors just mentioned, type of base and high/low F3 as factors; that is, the statistical analysis was conducted on "g" responses (or equivalently, "d" responses), not on percent correct. In this analysis, all effects with respect to percent correct are interactions involving the high/low F3 factor.

The first result evident from Figure 2 is that SOA had a clear effect: Performance decreased as SOA increased in either direction, $F(10,110) = 32.07$, $p < .0001$. A second clear effect is that of type of F3 segment: Performance was generally best for the dynamic F3 segments and poorest for the short F3 segments, $F(2,22) = 11.02$, $p < .0005$. Performance for the short F3 segments at SOA=0 in the dichotic condition was a good deal worse than in Experiment 1, for reasons that are not obvious. The third main effect evident from the figure is that, unexpectedly, performance in the diotic condition was higher than in the dichotic condition, $F(1,11) = 7.06$, $p < .03$.

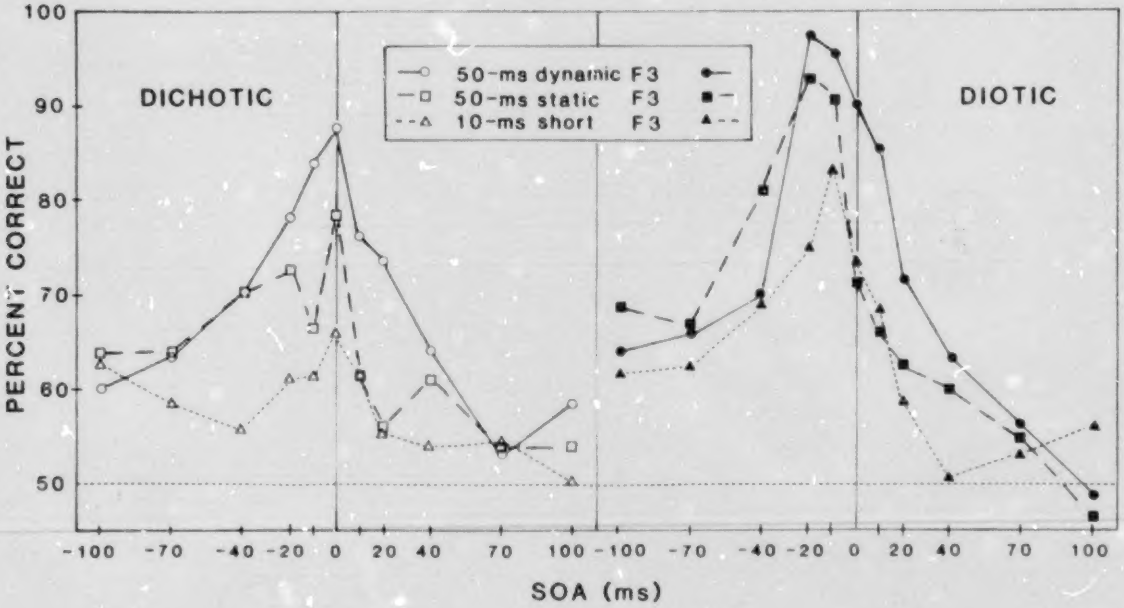


Figure 2. Percent correct as a function of SOA, separately for dichotic and diotic conditions, with type of F3 segment as parameter.

Because of the general convergence of scores at the extremes of the SOA range, interactions with SOA also reflect main effects, at least in part. These interactions were highly significant for both type of F3 segment, $F(20,220) = 5.51$, $p < .0001$, and presentation condition, $F(10,110) = 8.22$, $p < .0001$. Despite this latter interaction, listeners' tolerance of SOAs seemed similar in the two presentation conditions. No other effects on percent correct were significant.

Some more detailed differences in Figure 2 are not directly captured by the statistical analysis but deserve attention. First, in the dichotic condition performance was generally best at SOA=0, as expected, but in the diotic condition, optimal performance was at short negative SOAs. Second, the effect of SOA was generally asymmetric, though more so in the diotic than in the dichotic condition: Performance was generally better when the F3 segment led the base than when it lagged behind. This was especially true for the longest intervals used: At -70 and -100 ms of SOA, performance was clearly above chance ($p < .05$ for 11 of 12 conditions by sign test), whereas scores were near chance at 70 and 100 ms of SOA ($p < .05$ for only 1 of 12 conditions). Indeed, the absence of any decline in performance between -70 and -100 ms of SOA suggests an asymptote that may reflect an effect other than spectral/temporal fusion, such as a response bias contingent on the perceived pitch of the F3 segment. Third, it may be noted that the superiority of dynamic over static F3 segments did not hold at lead times of -40 ms or more, and that the superiority of static over short F3 segments was much more pronounced at negative than at positive SOAs.

One consequence of the differential asymmetry of the effect of SOA in the dichotic and diotic conditions is that diotic performance exceeded dichotic performance primarily at short F3 segment lead times. This is especially clear from Figure 3, where the difference between diotic and dichotic scores is plotted. It is also evident that this difference is similar for all three types of F3 segments. (The relevant interaction was not significant.)

The statistical analysis revealed several additional effects that related specifically to the percentage of "g" (or "d") responses, rather than to percent correct. Figure 4 shows the percentage of "g" responses as a function of SOA, high/low F3, and type of base; the scores are averaged over the three types of F3 segment and the two presentation conditions. Naturally, there were more "g" responses to stimuli including the low F3 than to stimuli including the high F3, $F(1,11) = 166.84$, $p < .0001$. It is also evident that the effect of the low F3 segment, which increased "g" responses when effective, was larger than that of the high F3, which decreased "g" responses, so that the total number of "g" responses varied significantly with SOA, $F(10,110) = 5.31$, $p < .0001$. Of course, the interaction of high/low F3 and SOA was highly significant; it corresponds to the main effect of SOA on percent correct, reported above. It may also be noted that the asymmetry around SOA=0 at short SOAs, deriving mainly from the diotic condition (cf. Figure 2), was pronounced only for low-F3 stimuli; the effect of SOA for high-F3 stimuli was more nearly symmetric. The asymmetry at long SOAs was equally present for both types of stimuli, however.

An unexpected result evident in Figure 4 is that, overall, more "g" responses were given when the base contained a steady-state F3, $F(1,11) = 17.13$, $p < .002$. The presence of a steady-state F3 apparently enhanced the spread of energy following the release, which is characteristic of velar consonants preceding back vowels. This difference was more pronounced at long than at short SOAs--$F(10,110) = 7.16$, $p < .0001$, for the interaction--which confirms that the effect originated in the base. However, the effect also interacted with type of F3 segment, $F(2,22) = 10.18$, $p < .0007$, being strongest with the short F3 segments and weakest with the dynamic F3 segments. Thus, the most effective F3 segments also were able to overcome most effectively the bias inherent in the base itself. A triple interaction between type of presentation, SOA, and high/low F3 was also obtained, $F(10,110) = 3.44$, $p < .0006$, suggesting that the bias was overcome more effectively by the F3 segments in the diotic condition. The differential SOA asymmetry in the two presentation conditions may also have contributed to this interaction.

Three additional significant interactions in the analysis of variance (between mode of presentation and high/low F3, between type of F3 segment and high/low F3, and between mode of presentation, type of F3 segment, and SOA) essentially parallel effects on percent correct described earlier and therefore need not be discussed any further.

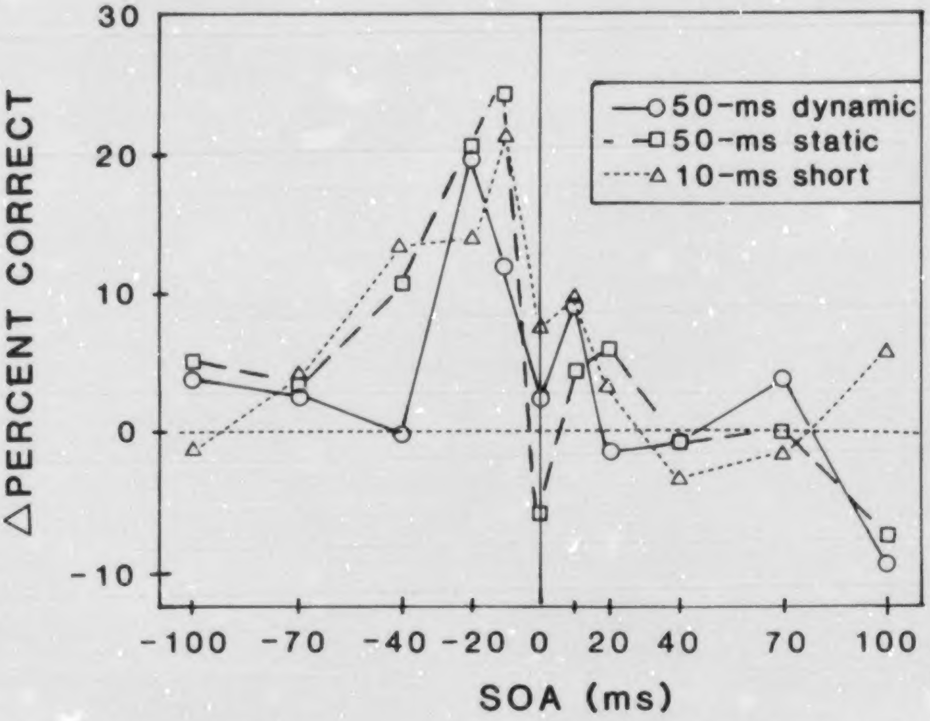


Figure 3. Difference between diotic and dichotic scores (Figure 2) as a function of SOA.

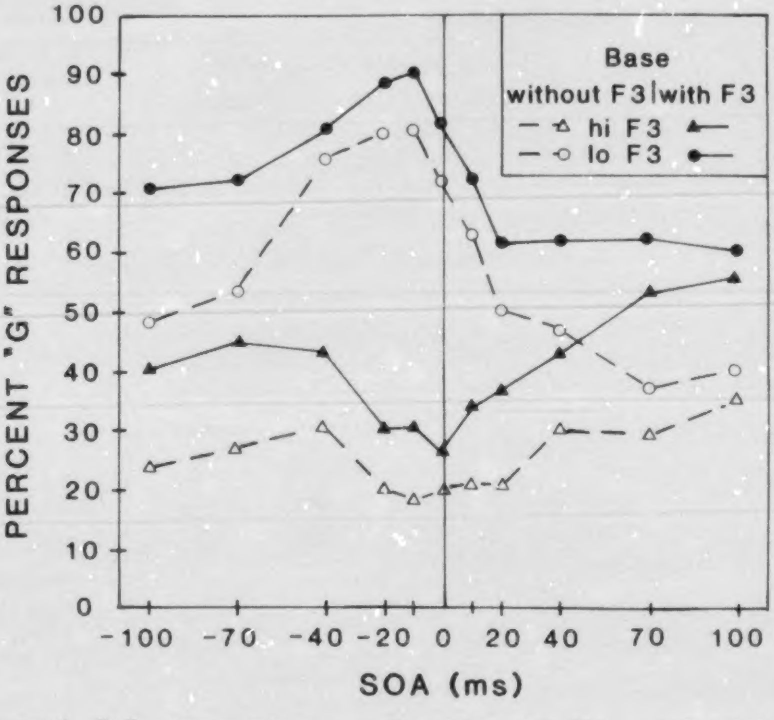


Figure 4. Percent "g" responses as a function of SOA, with high/low F3 and type of base (with or without F3 steady-state) as parameters.

Discussion

Experiment 2, in conjunction with Experiment 1, investigated three factors that were expected to play a role in spectral/temporal fusion of speech stimuli: (1) Structural properties of the isolated formant transition and of the base; (2) temporal asynchrony between the transition and the base; and (3) dichotic versus diotic presentation.

It is now clear that the isolated transition need not actually be a transition for fusion to occur. A steady-state formant with the same onset frequency, or even only the first pitch pulse of the transition can be sufficient, although the dynamic frequency transition does seem to convey additional information. Moreover, the base need not contain any continuation of the isolated F3 segment in the form of a steady-state F3. Experiment 2 has also shown that these same stimulus conditions enable listeners to discriminate /da/ and /ga/ in diotic presentation, when (at SOA=0) the stimulus components are physically integrated and the F3 segment is not perceived as a separate nonspeech stimulus. What is different about the dichotic situation is the presence of the added nonspeech percept: Segregation by input channel is effective at an auditory level of perception but apparently leaves phonetic perception unaffected, at least in the present paradigm.

This conclusion is also supported by the finding that the range of SOAs over which above-chance speech discrimination was obtained was very similar in dichotic and diotic presentation. Thus, even when the isolated F3 segment preceded the base on the same channel, it was nevertheless (partially) integrated with the base into a phonetic percept. Thus, the expectation that listeners would be less tolerant of SOAs in diotic presentation was not borne out, and the present results in fact suggest that spectral/temporal fusion is not specific to dichotic presentation at all. Nor is duplex perception: The F3 segment preceding the base on the same channel is perceived as a nonspeech event--a case of monaural duplex perception. We conclude that perceptual integration in phonetic perception operates regardless of mode of stimulus presentation, and apparently regardless of whether the stimulus appears unitary or segregated at an auditory level of perception. Although there are some obvious limits to this dissociation, it nevertheless strengthens further the traditional distinction between speech and nonspeech modes of perception.

There were two kinds of asymmetries with respect to the effects of SOA. One of them was equally present in dichotic and diotic presentation: Speech discrimination was above chance at long negative SOAs but dropped to chance at long positive SOAs. No such asymmetry was noted by Cutting (1976); however, the above-chance scores at long negative SOAs replicate the findings of Bentin and Mann (1983). Some of this asymmetry may be due to (central) masking of lagging F3 segments by the overlapping vowel; however, it seems that the above-chance performance with leading F3 segments is the finding in need of explanation. Only speculation is possible at this time. One possibility is that leading F3 segments are preserved in a (central) auditory memory and subsequently integrated with the base, whereas lagging F3 segments somehow cannot take advantage of auditory memory for the acoustically more complex base. Alternatively, identification of the F3 segment as "high" or "low" may have exerted a bias on speech identification, which was more pronounced when the F3 segment led than when it lagged the base. This explanation seems plausible, especially since the subjects were told about the correspondence of F3 segment pitch and phonetic category. Although they were also told to pay attention to the speech percept only, a certain amount of involuntary bias may have been introduced by leading F3 segments. This bias was equally present in diotic and dichotic presentation. Assuming, therefore, that the above-chance performance at long negative SOAs was not due to spectral/temporal fusion proper, the range of SOAs over which this type of fusion operates seems rather limited--roughly, ±50 ms.

The other asymmetry is the unexpected finding of optimal diotic performance at short negative SOAs. This was also the region where diotic performance exceeded dichotic performance. The following explanation may be proposed: Diotic integration of the stimulus components may have been uniformly superior to dichotic integration, but at positive SOAs diotic performance may have been lowered due to peripheral masking of the F3 segment by the lower formants contained in the base. Rand (1974) and many subsequent studies have suggested that dichotic segregation of a higher formant from F1 results in a release from upward spread of masking, which thus is largely a peripheral (channel-specific) effect. In fact, it was surprising that the present data did not show an absolute advantage for dichotic presentation at SOA=0 and at positive SOAs. The upward spread of masking explanation may account for another feature of the present data that seems difficult to explain in other terms: Apparently, the asymmetry in the diotic SOA effect was entirely due to the low F3; stimuli with a high F3 showed no such asymmetry. The reason for this may be that the high F3 evaded masking by the F1 and F2 transitions. The present data thus seem consistent with earlier findings on upward spread of masking, if the assumption is granted that dichotic fusion was not quite as strong as in some of the earlier studies.

An alternative possibility that comes to mind is that an F3 segment protruding from the base (at short negative SOAs) may have been perceived as if it were a release burst. This would explain why speech identification was more accurate at short F3 lead times than at lag times, but it would not be clear why this asymmetry was present only in the diotic condition and only for the high-pitched F3. Nor did the 50-ms F3 segments sound like noisy release bursts; they had a distinct tonal quality. Thus, without additional assumptions yet to be spelled out, this interpretation cannot account for the data.

In summary, the present findings reveal dichotic spectral/temporal fusion to be a phenomenon that is neither specifically dichotic nor specifically temporal. The fact that a temporally or spatially segregated formant segment is audible as a separate nonspeech sound is not surprising; that such an auditorily segregated stimulus component still contributes to an integrated phonetic percept, however, is an observation that deserves continued attention. Although Pastore, Schmuckler, Rosenblum, and Szczesiul (1983) have reported a somewhat analogous phenomenon with musical stimuli, it is still possible to entertain the hypothesis that the fusion effect studied here reflects the operation of a central integrative mechanism specialized for phonetic perception. This hypothesis needs to be tested further with nonspeech analogs of speech stimuli used in studies of spectral/temporal fusion.

References

Bentin, S., & Mann, V. A. (1983). Selective effects of masking on speech and nonspeech in the duplex perception paradigm. Haskins Laboratories Status Report on Speech Research, SR-76, 65-85.

Broadbent, D. E. (1955). A note on binaural fusion. *Quarterly Journal of Experimental Psychology*, *7*, 46-47.

Broadbent, D. E., & Ladefoged, P. (1957). On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, *29*, 708-710.

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review*, *83*, 114-140.

Danaher, E. M., & Pickett, J. M. (1975). Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss. *Journal of Speech and Hearing Research*, *18*, 261-271.

Darwin, C. J., Howell, P., & Brady, S. A. (1978). Laterality and localization: A right ear advantage for speech heard on the left. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, NJ: Erlbaum.

Liberman, A. M. (1979). Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In E. Fischer-Jørgensen, J. Rischel, & N. Thorsen (Eds.), *Proceedings of the IXth International Congress of Phonetic Sciences* (Vol. II, p. 468). Copenhagen: University of Copenhagen.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist*, *37*, 148-167.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, *30*, 133-143.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, *14*, 211-235.

Nábělek, I. V., Nábělek, A. K., & Hirsh, I. J. (1970). Pitch of tone bursts of changing frequency. *Journal of the Acoustical Society of America*, *48*, 536-553.

Nearey, T. M., & Levitt, A. G. (1974). Evidence for spectral fusion in dichotic release from upward spread of masking. *Haskins Laboratories Status Report on Speech Research*, *SR-39/40*, 81-89.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1983). The role of "chirp" identification in duplex perception. *Perception & Psychophysics*, *33*, 323-332.

Nye, P. W., Nearey, T. M., & Rand, T. C. (1974). Dichotic release from masking: Further results from studies with synthetic speech stimuli. *Haskins Laboratories Status Report on Speech Research*, *SR-37/38*, 123-137.

Pastore, R. E., Schmuckler, M. A., Rosenblum, L., & Szczesiul, R. (1983). Duplex perception with musical stimuli. *Perception & Psychophysics*, *33*, 469-474.

Pastore, R. E., Szczesiul, R., Rosenblum, L. D., & Schmuckler, M. A. (1982). When is a [p] a [t], and when is it not. *Journal of the Acoustical Society of America*, *72* (Supplement No. 1), S16 (Abstract).

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, *55*, 678-680.

Repp, B. H., Milburn, C., & Ashkenas, J. (1983). Duplex perception: Confirmation of fusion. *Perception & Psychophysics*, *33*, 333-337.

Schwab, E. C. (1981). *Auditory and phonetic processing for tone analogs of speech*. Unpublished doctoral dissertation, SUNY at Buffalo.

MONITORING FOR VOWELS IN ISOLATION AND IN A CONSONANTAL CONTEXT*

Brad Rakerd,† Robert R. Verbrugge,†† and Donald P. Shankweiler†††

Abstract. The identifiability of isolated vowels (/V/) was compared to that of vowels in consonantal context (/pVp/) when subjects performed a monitoring task. On successive blocks of trials in a test series, the subjects listened for instances of one or another of nine monophthongal vowels (/i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,ʊ,u/) and identified each test item as being an instance or not. On average, resulting false alarm errors occurred significantly less often in the /pVp/ condition, consistent with the previous finding that vowel perception may be aided by consonantal context. This beneficial effect of context was found to be restricted to the class of open vowels, however, with perception of the close vowels being somewhat hindered by context. The error data for misses also showed an interaction between context and vowel height. Various accounts of interaction are considered.

Of continuing interest in speech research is the question of whether vowel perception is affected by the consonantal context in which a vowel occurs. Perceivers might be expected to exhibit some context sensitivity because the acoustic correlates of a vowel often vary with changes in the identity of neighboring consonants (Broad, 1976; Lindblom, 1963; Stevens & House, 1963). Strong support for this hypothesis comes from studies in which vowels have been shown to be more identifiable in a consonantal context than in isolation (e.g., Gottfried & Strange, 1980; Strange, Edman, & Jenkins, 1979; Strange, Verbrugge, Shankweiler, & Edman, 1976).

Recently, this evidence has been challenged on grounds that it is largely an artifact of the perceptual task subjects have been asked to perform. It has typically been required that subjects make a multiple-choice identification judgment by: (1) selecting the "best match" to a presented vowel from among a prescribed set of alternatives; and (2) indicating their choice by circling a written form of the alternative on an answer sheet. That written form can be orthographically related to a presented item in varying degrees. It can, for example, be an English spelling of the item itself (e.g., "pep" as the correct response to /pɛp/), or it can be a spelling of a word that con-

---

*Journal of the Acoustical Society of America, 1984, 76, 27-31.

†Department of Psychology, Michigan State University.

††Bell Laboratories.

†††Also University of Connecticut.

Acknowledgment. This research was supported by NICHD grant HD-01994 and BRS RR-05596 to Haskins Laboratories. We wish to thank Winifred Strange and Carol Fowler for their assistance in the design of this study and for their comments on earlier versions of the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-79/80 (1984)]

tains the "same" vowel as the presented item (e.g., "bed" as the correct response for /pεp/). The degree of relationship between item and response alternative has been shown to affect vowel identification performance (Assmann, Neary, & Hogan, 1982; Diehl, McCusker, & Chapman, 1981; Macchi, 1980). This variable was not controlled in early studies of consonantal context (e.g., Strange et al., 1976), therefore the significance of the obtained effect has been called into question (but see Strange & Gottfried, 1980).

The significance of the context effect has also been questioned on the argument that the typical response task--i.e., the searching for and circling of an appropriate alternative on an answer sheet--is itself somewhat biased in favor of the context condition. This is owing to the fact that such a task makes strong demands on short-term memory in that a stimulus trace must be held long enough to be compared with each of the alternatives. Vowels in context might be expected to be somewhat better remembered than isolated vowels for two reasons: (1) vowel-consonant combinations tend to make up words already represented in a subject's lexicon, or at least portions of such words; and (2) in English, the orthographic representations of vowels in context tend to be less ambiguous than those of isolated vowels (see Diehl et al., 1981, for elaboration on this argument).

In light of these methodological concerns over past work, we thought it useful to make a comparison of the identifiability of vowels in and out of context with a different kind of perceptual task than has previously been employed. While such a task would, no doubt, have certain limitations of its own, it was felt that if these were sufficiently different from the limitations of the multiple-choice identification task, the results could speak to the methodological generalization of any effects of consonantal context. The specific task we set for subjects was that of monitoring lists of test items for instances of particular target vowels. Subjects simply checked "yes" on an answer sheet if a presented item (an isolated vowel or a vowel in context) was judged to be an instance of the target vowel being monitored and "no" if it was not.

This method has two virtues that are noteworthy: it is comparatively free from orthographic bias since there are no written vowel alternatives on the answer sheet, and it imposes minimal memory demands on the subject since a presented item can be immediately judged to match the target or not. Monitoring thus affords a good comparison with the identification method of past studies. Here, we strengthened that comparison further by examining vowel stimuli for which perceptual data had already been collected with the previous method (Strange et al., 1976).[1]

Experimental Methods

Stimuli

All /pVp/ and /V/ stimuli were produced by a single male talker who spoke an Upper Midwestern dialect of English. For each condition, he produced five tokens of each of the nine vowels. These were organized into /pVp/ and /V/ test series according to the following protocol: (1) 90 items (two repetitions of each token) were assembled in randomized order to make up a block; (2) monitoring instructions identifying the particular vowel to be listened for in that block were inserted at its beginning; (3) instructions reminding the subject of the target vowel were inserted after the 30th and 60th items;

(4) steps (1) through (3) were repeated for a total of nine test blocks. There was a 2-s pause between test items and a 30-s pause between blocks.

The monitoring and reminder instructions were recorded by a male speaker with the same dialect as that of the speaker who had produced the test stimuli. For both experimental conditions, the monitoring instructions were given in the following form: "In this test block, you will be listening for the vowel (exemplar 1), as in (CVC 1), (CVC 2), (CVC 3). Listen for the vowel (exemplar 2), (exemplar 3), (exemplar 4)." The exemplars were isolated productions of the vowel. The CVCs were English monosyllabic words that contained the vowel.[2] The reminder instructions were as follows: "Remember, you are listening for the vowel (exemplar 5), (exemplar 6), (exemplar 7)."

The order in which vowels were monitored was varied across listeners; nine different orders were generated with the constraint that each of the nine vowels was monitored in each ordinal position.

Acoustic characteristics of the stimuli. These stimuli are a subset of the items employed in a previous study of vowel perception (Strange et al., 1976). Their acoustic characteristics conform to generalizations reported in that study. The first of these generalizations is that the formant frequencies of all isolated vowels except /ɔ/ were comparable to normative values reported by Peterson and Barney (1952). The deviations in /ɔ/ reflect an idiosyncrasy of the speaker's dialect. Average first formant frequencies for the vowels in /pVp/ context were comparble to the values for isolated vowels. In contrast, the second formant frequencies of /pVp/ vowels were somewhat "reduced" (cf. Lindblom, 1963) relative to the isolated vowels. That is to say, they exhibited a somewhat smaller range of deviation about the average value for all vowels in the set.

The isolated vowels were, on average, considerably longer than /pVp/ vowels. Relative durations of vowels in the two conditions were roughly comparable, however. As might be expected on the basis of previous reports (e.g., Peterson & Lehiste, 1960), the vowels /ɪ,ɛ,ʌ,ʊ/ generally were the briefest in duration, /i,u/ were intermediate, and /æ,ɑ,ɔ/ were the longest. The exceptions to this were the vowel /u/ in the /pVp/ context and the vowels /ɑ,ɔ/ in isolation, all of which were somewhat shorter than expected.

Subjects

Thirty-six undergraduates enrolled in an introductory psychology course at the University of Connecticut, participated in this experiment. They were randomly assigned to the /pVp/ and /V/ conditions, so that there were 18 subjects in each condition. All of the subjects were adult native speakers of English with normal hearing. They had no knowledge of the purpose of this study.

Procedure

Subjects were asked to monitor the lists of test items for occurrences of the monophthongal vowels /i,ɪ,ɛ,æ,ʌ,ɑ,ɔ,ʊ,u/. They reported their decisions by checking "yes" on an answer sheet if an item was judged to be an instance of the vowel being monitored on a trial and "no" if it was not.

Instructions and test materials were presented over headphones with the volume adjusted to a comfortable listening level, conditions comparable to those employed in the previous identification study conducted with these same stimuli (Strange et al., 1976). Subjects were tested, two at a time, in a sound-attenuated room. Before the start of testing, they were familiarized with the stimulus and response materials in the following way: First, the testing procedure was described. It was explained that a number of different speech stimuli would be presented and that the task would be to monitor the vowels in the manner described above. Next, a randomly selected sample of the stimuli was presented. For the first few trials (approximately 15), subjects were asked to listen to the stimuli and make no response. Then, they were given a sample answer sheet and, for 30 trials, monitored the sample items for instances of a particular target vowel. This target was randomly varied across subjects. No feedback was given as to the accuracy of these practice responses; subjects were, however, allowed to ask questions of clarification about all aspects of the procedure. The test was begun only after all subjects expressed confidence that they completely understood the task.

Results

With this monitoring procedure, subjects could make errors of two types: false alarms and misses. False alarms were erroneous acceptances of vowels other than the target--responding "yes" when the correct choice was "no." Misses were failures to recognize actual instances of the vowel being monitored--responding "no" when the correct choice was "yes." Neither type of error was significantly related to the order in which the vowels were monitored; consequently, the data that will now be considered were pooled across monitoring orders.

False Alarms

In the left half of Table 1, *composite* false alarm error rates are summarized for each vowel category. A composite false alarm resulted whenever a presented vowel was erroneously taken to be an instance of any of the other alternatives. For example, in the isolated condition, listeners variously misheard the vowel /ʌ/ to be an instance of /æ/, /ɑ/, and /ʊ/. Together, these false alarms occurred on 8.4% of all trials in which /ʌ/ was the presented vowel but was not, in fact, the target. Since many vowel pairs (/ʌ-i/ for instance) were seldom if ever confused, averaging over all of the alternatives in this way generally resulted in rather low error rates. However, this measure of false alarms is perhaps the most comparable to the miss percentage to be considered below and it will be seen that the data exhibit a similar structure.

The two leftmost columns of Table 1 report composite false alarm rates for the consonantal-context (/pVp/) and isolated (/V/) conditions, respectively. The difference in error rates between these two conditions is given in the third column (/pVp/-/V/). Results for the vowel /ɔ/ are reported separately in the table. This is because the acoustic characteristics of /ɔ/ proved to be abnormal and because this vowel behaved differently than the other open vowels, both here and in the comparison study of Strange et al. (1976). (For further consideration of this difference see the Discussion section below.) Arc sin transformations of the composite false alarm data shown in Table 1, and of all other data to be discussed, were submitted to analysis of variance.[3]

Table 1

Average Composite False Alarm and Miss Error Rates

Percentage Errors

	Composite False Alarms			Misses		
Vowel	/pVp/	/V/	/pVp/-/V/	/pVp/	/V/	/pVp/-/V/
i	.5	.2	+.3	2.2	1.7	+.5
ɪ	.3	1.9	-1.6	6.1	1.1	+5.0
ɛ	1.7	7.7	-6.0	6.1	11.1	-5.0
æ	1.2	2.4	-1.2	3.9	11.7	-7.8
ʌ	3.3	8.4	-5.1	6.1	17.8	-11.7
ɑ	4.0	5.2	-1.2	26.1	41.1	-15.0
ʊ	4.7	2.7	+2.0	17.8	13.3	+4.5
u	2.2	.6	+1.6	6.7	.6	+6.1
Overall	2.2	3.6	-1.4	9.4	12.3	-2.9
/ɔ/	9.2	6.0	+3.2	9.4	3.9	+5.5

Table 2

Average Error Rates for the Major False Alarm Vowel Pairs

	Percentage of False Alarm Errors		
Vowel Pair	/pVp/	/V/	/pVp/-/V/
/ɛ/-/æ/	4.8	30.8	-26.0
/ʌ/-/ɑ/	14.2	31.6	-17.4
/ʌ/-/ʊ/	15.6	13.0	+2.6
/ʊ/-/u/	9.5	3.6	+5.9
Overall	11.0	19.8	-8.8
/ɔ/-/ɑ/	50.8	59.7	-8.9

Two of the findings regarding composite false alarms speak to the question of whether or not consonants exert a contextual influence on vowel perception. The first is that, overall, error rates in the consonantal condition were significantly lower than in the isolated condition, $F(1,34) = 4.20$, $p < .05$. This indicates that when listeners monitor vowels, as when they perform other identification tasks (Gottfried & Strange, 1980; Strange et al., 1979; Strange et al., 1976), their performance may be positively influenced by the presence of neighboring consonants. The second finding is that the beneficial effect of context was not in evidence for all vowels (see column three of Table 1). Generally speaking, it was the perception of open vowels that was aided by context, with perception of close vowels proving to be somewhat poorer in the context condition. The only exceptions to this generalization were the vowels /ɔ/, which behaved anomalously throughout, and /ɪ/, which was seldom confused with the other vowels in either condition. This difference between the open and close vowels was reflected in a significant interaction between context and vowel height, $F(1,34) = 20.84$, $p < .001$. Post hoc examination of this interaction revealed that the simple main effect of context was significant only for open vowels, $F(1,34) = 18.20$, $p < .001$.

As noted above, false alarm errors occurred only rarely for many of the vowel pairs. However, a few pairs did show false alarm rates that were rather high. These are summarized in Table 2. Note that the mean false alarm rate for these vowel pairs was at least five times as great as the mean composite false alarm rate in both the /pVp/ and /V/ conditions. Hence, these high-likelihood false alarm pairs were the major contributors to overall error scores. The two observations made about the composite false alarm data apply to these high-likelihood false alarms as well. First, overall identifiability of the vowels was enhanced by context. There were significantly fewer errors in the /pVp/ condition, $F(1,34) = 8.88$, $p < .01$. Second, there was a significant interaction between context and vowel height, $F(3,102) = 11.05$, $p < .001$, reflecting the fact that errors on open vowel pairs occurred significantly less often in consonantal context,[4] /ɛ-æ/: $F(3,102) = 25.64$, $p < .001$; /ʌ-ɑ/: $F(3,102) = 13.62$, $p < .001$, and those on the close pair (/ʊ-u/) occurred more often but not significantly so.

Misses

Miss errors are reported on the right half of Table 1. It can be seen that their overall pattern parallels that of false alarms. Subjects were, however, much more variable in exhibiting the pattern with misses. As a consequence, the main effect of context was not significant for these data, $F(1,34) < 1.0$. There was a highly significant context-by-vowel height interaction, $F(1,34) = 15.54$, $p < .001$. As before, this resulted from the fact that performance on the open vowels was significantly aided by context, $F(1,34) = 8.90$, $p < .01$, while that on the close vowels was hindered to a lesser and nonsignificant degree. Also as before, /ɔ/ behaved differently from the other open vowels. It was missed more frequently in the consonantal condition than in isolation.

The Question of Response Biases

Although we have looked at false alarm and miss errors separately, the two are not strictly independent. Notice, for example, that if the subjects in this experiment had (for any reason) chosen to respond "yes" on all monitoring trials, we would have observed no miss errors and 100% false alarms.

Conversely, a bias toward "no" responses would have inflated misses and deflated false alarms. It is therefore reasonable to wonder whether some or all or the effects that we observed can be attributed to systematic response biases. Given the overall patterning of the two types of errors, this possibility can be confidently rejected. It has been noted throughout that the data structure of the miss and false alarm errors was roughly the same. If there were significant response biases, we should have expected the two types of errors to have complementary distributions, not comparable ones. For example, we should have expected that the observed interactions between context and vowel height would have been in opposite directions for the two types of errors. They were not.

Discussion

We compared listeners' ability to identify vowels in and out of a consonantal context (/pVp/) when they performed a monitoring task and found that they made significantly fewer false alarm errors (both composite false alarms and high-likelihood false alarms) in the /pVp/ condition. This clearly supports the view that the contextual advantage for vowel perception observed here and elsewhere (Gottfried & Strange, 1980; Strange et al., 1979; Strange et al., 1976) is a genuine perceptual effect and not simply a methodological artifact. At the same time, however, these monitoring results add to evidence indicating that the demonstrability of a contextual influence may be greatly affected by task variables. Pooling misses and false alarms, our subjects made an average of 4.1% errors in the /pVp/ condition and 5.3% in the /V/ condition. In the comparison identification study conducted with these same stimuli (Strange et al., 1976), substantially different error rates were reported. In that instance, there were 9.7% errors in the /pVp/ condition and 33.1% in the /V/ condition. Clearly, absolute error rates can vary substantially with the method of assessment, and these form the baseline against which any relative influence of consonantal context must be measured.

There were two additional points of agreement with the study of Strange et al. (1976) that merit comment. The first involves the vowel /ɔ/, which did not behave like the other open vowels in the present instance. It turns out that perception of /ɔ/ was anomalous in that earlier study as well. This can be seen in Table 3, which summarizes their multiple-choice identification data for our speaker's tokens (these data are excerpted from the segregated-talker condition of Experiment I in Strange et al., 1976). Notice that with their method Strange et al. observed a contextual advantage for the identification of all vowels except /ɔ/. It appears that the unusual perception of this vowel reflects some abnormality in its production. This conclusion is further supported by the fact that formant frequencies for /ɔ/--as produced in both conditions--were very different from population norms.

The second point of comparison with Strange et al. (1976) concerns the perceptual interaction between context and vowel height that we observed. Some analog to that interaction can also be seen in their data. Note in Table 3 that while all vowels in their /pVp/ condition (except /ɔ/) were identified more accurately than the isolated counterparts, open vowels were much more aided by context than the close vowels. The mean contextual advantage (/pVp/-/V/) for the open vowels was 40.3%, while that for the close vowels was only 12.7%. In both studies, then, we see some evidence that the presence of a /pVp/ context differentially affected perception of the open and close vowels.

Table 3

Identification Error Rates Determined by Strange et al. (1976). Data are for the Single Male Talker in Their Segregated-talker Condition of Experiment I.

	Percentage of Identification Errors		
Vowel	/pVp/	/V/	/pVp/-/V/
i	0.0	11.0	-11.0
ɪ	0.9	14.0	-13.1
ɛ	1.8	63.0	-61.2
æ	1.8	19.0	-17.2
ʌ	6.4	57.0	-50.6
ɑ	42.7	75.0	-32.3
ʊ	15.5	33.0	-17.5
u	1.8	11.0	-9.2
ɔ	16.4	15.0	+1.4
Overall	9.7	33.1	-23.4

Though the acoustic and/or articulatory origins of this effect are yet to be confidently determined, we can make some preliminary observations. First, we may note that no satisfactory explanation of it is likely to be advanced in terms of formant frequency differences among the vowels. Owing to the phenomenon of vowel reduction (Lindblom, 1963), those differences were in fact less great in the more perceptually distinctive /pVp/ condition. A more promising acoustic account is that the perceptual effect somehow results from the greater degree of spectral change associated with open vowels. In /pVp/ context, open vowels are typically marked by more extensive formant transitions out of and into the flanking consonants than are close vowels. Vowel height should be particularly related to transitions of the first formant. There have been speculations that acoustic dynamics of this sort positively influence vowel perception (Strange, Jenkins, & Johnson, 1983; Strange et al., 1976).

The acoustics also provide evidence that vowels and consonants were coarticulated in the /pVp/ condition--vowel formant frequencies were reduced in this context. This has led us to consider an articulatory account of the perceptual effect. It may be that the beneficial influence of /pVp/ context was focused on the open vowels because those vowels are coarticulated with the consonants in some manner in which close vowels are not. This could pertain particularly to articulatory movements of the jaw. The jaw lowering required for production of open vowels must be coordinated with jaw raising to achieve bilabial closure for the consonants. While production of the close vowels would likewise call for some jaw lowering (and hence for some articulatory coordination with the consonants), it is conceivable that this requirement differs in kind or degree from that for the open vowels. If listeners are aware of such a coarticulatory difference, it could affect their interpretation of the acoustic signal.

We plan to distinguish between these alternative accounts of the perceptual effect by looking at vowel monitoring performance in other consonantal contexts. While perception of the open vowels was particularly aided by /pVp/ context in the present study, we expect that rather different interactions will occur with consonants of some other place and manner of articulation.

References

Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America*, *71*, 975-989.

Broad, D. J. (1976). Toward defining acoustic phonetic equivalence for vowels. *Phonetica*, *33*, 401-424.

Diehl, R. L., McCusker, S. B., & Chapman, L. S. (1981). Perceiving vowels in isolation and in consonantal context. *Journal of the Acoustical Society of America*, *68*, 239-248.

Gottfried, T. L., & Strange, W. (1980). Identification of coarticulated vowels. *Journal of the Acoustical Society of America*, *68*, 1626-1635.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, *35*, 1773-1781.

Macchi, M. J. (1980). Identification of vowels spoken in isolation and in consonantal context. *Journal of the Acoustical Society of America*, *68*, 1636-1642.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175-184.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, *6*, 111-128.

Strange, W., Edman, T. R., & Jenkins, J. J. (1979). Acoustic and phonological factors in vowel identification. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 643-656.

Strange, W., & Gottfried, T. L. (1980). Task variables in the study of vowel perception. *Journal of the Acoustical Society of America*, *68*, 1622-1625.

Strange, W., Jenkins, J. J., & Johnson, T. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, *74*, 695-705.

Strange, W., Verbrugge, R., Shankweiler, D. P., & Edman, T. R. (1976). Consonantal environment specifies vowel identity. *Journal of the Acoustical Society of America*, *60*, 213-224.

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw Hill.

Footnotes

[1] Those earlier data are for the single male talker in the segregated-talker condition of Experiment I in Strange et al. (1976).

[2] The CVCs in the monitoring instructions were as follows. For the vowel /i/: *green*, *peak*, *seal*; /ɪ/: *bit*, *tin*, *sick*; /ɛ/: *pen*, *wet*, *step*; /æ/: *hat*, *fan*, *map*; /ʌ/: *cup*, *gum*, *rut*; /ɑ/: *top*, *sock*, *dot*; /ɔ/: *fog*, *call*, *gone*; /ʊ/: *put*, *look*, *should*; /u/: *boot*, *cool*, *moon*.

[3] Because the error rates were often close to zero, they were transformed according to the following formula suggested by Winer (1962):

$$X' = 2 \text{ arc sin } \sqrt{X + 1/N}$$

Where X is the original score, X' is the transformed score, and N is the number of subjects in a condition.

A one-within (vowel height), one-between (context) analysis of variance was performed on the transformed scores.

[4] In this instance /ɔ/ again proved to behave differently from the other open vowels. False-alarm confusions between members of the vowel pairs /ɛ-æ/ and /ʌ-ɑ/ were greatly reduced by /pVp/ context for both "directions"--i.e., with regard to confusions of the first member of the pair with the second and the second with the first. This was not the case with /ɔ-ɑ/, however. In the /pVp/ condition, /ɑ/ was misheard as /ɔ/ much less frequently than in isolation (47.2% false alarms vs. 75.5%), but /ɔ/ was misheard as /ɑ/ more frequently in this condition (54.4% false alarms vs. 43.9%). The data reported in Table 2 reflect the average of these two types of confusions.

PERCEPTION OF [l] AND [r] BY NATIVE SPEAKERS OF JAPANESE: A DISTINCTION BETWEEN ARTICULATORY AND PHONETIC PERCEPTION

Virginia A. Mann†

Abstract. Although native speakers of Japanese may be unable to identify the phonemes [l] and [r] in English, they, like native speakers of English, unconsciously take account of certain articulatory differences between these speech sounds. One implication is that, preceding a language-specific level of speech perception where utterances are represented in terms of their constituent phonemes, there may exist a universally-shared level of speech perception where utterances are represented as articulatory patterns.

What do native speakers of Japanese perceive as they listen to English utterances that contain [l] and [r]? In the absence of considerable experience with spoken English, many Japanese are unable to label, discriminate, or produce [l] and [r] in a consistent fashion (Goto, 1971; Miyawaki et al., 1975; Mochizuki, 1981), which would seem to suggest that they hear these two speech sounds as one and the same. This study offers evidence that whether or not Japanese subjects can identify [l] and [r] phonetically, they tacitly perceive an articulatory difference between these speech sounds.

To demonstrate that Japanese speakers can perceive an articulatory difference between [l] and [r], though not a phonological one, this study has focused on a specific context effect in speech perception (for a general discussion of such effects, see Repp, 1982). The effect occurs when utterances that end in [l] or [r] precede utterances that begin with [d] or [g]. It may be demonstrated by placing the spoken syllables [al] and [ar] in front of stimuli from along a continuum of synthetic speech syllables ranging from [da] to [ga]. The presence of the preceding syllables causes systematic shifts in the category boundary between [d] and [g]: When the preceding syllable is [al], the boundary is shifted towards more [g] percepts (less [d] percepts), relative to that obtained when the preceding syllable is [ar] (Mann, 1980).

---

†Also Department of Psychology, Bryn Mawr College, Bryn Mawr, PA.

Acknowledgment. This study was completed at the Research Institute of Logopedics and at the Komaba Campus of the University of Tokyo, while the author was a Fulbright Fellow. The study was partially supported by NICHD grant HD-01994 and BRS grant RR-05596 to Haskins Laboratories. Recognition is due to Dr. Shigeru Kiritani and Dr. Hiroshi Suzuki for their advice and for their help in procuring subjects and a testing site. Ms. Michiko Mochizuki-Sudo is to be thanked for translating the instructions to the subjects.

By using the phenomenon known as duplex perception (Liberman, Isenberg, & Rakerd, 1981; Rand, 1974), it has been possible to demonstrate that the context effect of [l] and [r] on perception of [d] and [g] is not due to some general property of acoustic perception, but is highly specific to the perception of speech (Mann & Liberman, 1983). In duplex perception, one and the same stimulus is simultaneously heard as speech and as nonspeech. This situation can be created by dividing synthetic speech syllables along a [da] to [ga] continuum into two parts: a constant base portion that tends to sound like [da], and a third formant transition that in isolation sounds like a "chirp," but when combined with the base provides the critical cue for the distinction between [da] and [ga]. When base and transition are presented dichotically, the third formant transition is simultaneously perceived in two ways: as speech and nonspeech. It provides critical support for the perception of [da] or [ga] but also for the nonspeech "chirp." Listeners can be instructed to attend to one or the other of these percepts, and under instructions to ignore the speech percepts and attend to the nonspeech chirps, perception is continuous, and no context effect occurs when stimuli are preceded by [al] or [ar]. In contrast, under instructions to label or discriminate stimuli on the basis of the speech percepts [da] and [ga], perception is categorical and the location of the category boundary can be manipulated by the presence of a preceding syllable [al] or [ar]. Thus the context effect of [al] and [ar] is evident only when the stimuli are perceived as speech.

The explanation of why, in speech perception, [l] and [r] alter the position of the [d]-[g] boundary, rests on two related observations. First, it has been found that the effect of a preceding consonant on the distinction between [da] and [ga] is not limited to [l] and [r], but extends to the fricatives, [s] and [ʃ] (Mann & Repp, 1981), and that similarities are better described in terms of articulatory, than auditory properties. Specifically, preceding [l] and [s], which are produced with the tongue relatively forward in the mouth, shift perception away from [da] toward the more backwards [ga], relative to preceding [r] and [ʃ], which are produced with a more retracted tongue posture. Second, it has been shown that the perceptual effects of [l] and [r] find a parallel in speech production, where, owing to coarticulation, the acoustic structure of [da] and [ga] can vary as a function of whether they follow [l] or [r] (Mann, 1980). Both observations support the view that the context effects of [l] and [r], along with many other context effects and trading relations (see, for example, Repp, 1982; Repp, Liberman, Eccardt, & Pesetsky, 1978), represent a perceptual sensitivity to the consequences of coarticulation in the speech signal. Human listeners appear to possess some tacit knowledge about articulation and its consequences on the speech signal, and application of that knowledge may be part of what makes speech perception "special" (see, for example: Best, Morrongiello, & Robson, 1981; Liberman, 1982; Mann & Liberman, 1983; Repp et al., 1978).

Aside from revealing the special nature of perception in the speech mode, studies of the context effect of [l] and [r] on perception of [da] and [ga] can offer insight into the relationship between articulation-based perceptual adjustment, phonetic perception, and specific language experiences if they compare native speakers of Japanese with those of English. English and Japanese share many phonetic types, including [d] and [g], but Japanese does not distinguish the liquids [l] and [r] (its single "liquid," [r], more clearly resembles an alveolar flap than English [ɹ]). Consequently, absence of early experience with this phonetic contrast renders many native speakers of

Japanese unable to distinguish English utterances that contain [l] and [r] in phonetic labeling tasks, discrimination tasks, and in their own productions (Goto, 1971; Miyawaki et al., 1975; Mochizuki, 1981). Yet two- to three-month-old American infants have been found capable of making some discrimination between utterances that contain [l] and [r] (Eimas, 1975), and the contrast raises questions about the nature of native endowment and the role of experience in the development of speech perception. The present context effect offers a means of answering some of these questions.

One explanation of the speech perception abilities of infants vis-a-vis the phonetic difficulties of native speakers of Japanese is that a lack of specific experience has led to a loss of all ability to perceive a difference between [l] and [r] (Eimas, 1975). Another, slightly different view holds that infants may not perceive [l] and [r] as different phonemes so much as they perceive them as different articulatory patterns. If so, lack of experience with the [l]-[r] distinction might lead to an inability to distinguish [l] and [r] phonetically, but not necessarily to a desensitization of the basic ability to apprehend the articulatory differences between them. Using the present context effect, one can test this possibility, by asking whether Japanese subjects who cannot phonetic categorize [l] and [r] can nonetheless take account of articulatory differences between them.

Method

Subjects

Sixteen college freshmen enrolled in the first semester of a spoken English course at the University of Tokyo participated in the study. All were native speakers of Japanese who had never lived in an English-speaking society. They were selected by the English professor from a population of 150 students, on the basis of either superior (N=8) or inferior (N=8) performance on two standardized tests of spoken English perception and comprehension. In addition to these native speakers of Japanese, the experiment further included a control group of ten native speakers of English. They were undergraduates attending Bryn Mawr and Haverford Colleges.

Procedure

The experiment was divided into three stages and employed materials that have been described in detail elsewhere (Mann, 1980): a seven-member synthetic [da]-[ga] continuum and 12 natural tokens of [al] and [ar]. Stimuli along the [da]-[ga] continuum comprised three-formant syllables in which systematic variations in the onset of the third formant provided critical support for the [d]-[g] distinction. They were constructed so as to be compatible with the natural tokens of [al] and [ar]. Those tokens had been extracted from natural productions by a male speaker of English of [al-da], [al-ga], [ar-da], and [ar-ga] in which the first syllable had been stressed. To control for the possibility of material-specific effects, three tokens of each production were used.

In the first stage of the experiment, isolated stimuli from along the [da]-[ga] continuum were presented 12 times each, according to a randomized sequence. In the second, the [da]-[ga] stimuli were preceded by the tokens of [al] and [ar] and again presented 12 times in each context, according to an unblocked randomized sequence. In each stage, a 28-item practice sequence of the test items preceded the test sequence itself, and the task was to mark (on a response sheet containing both alphabetic script and Japanese Kana) whether a given stimulus contained [da] or [ga]. The third and final stage assessed subjects' ability to identify [l] and [r] in the stimuli previously employed in the second stage of testing, by marking (on a response sheet written in alphabetic script) whether a given stimulus contained [al] or [ar]. In light of the potential difficulty of this task, listeners were first pretrained in the appropriate response categories for 28 items, and then given a practice sequence of 28 items in which they were told the correct response before listening to each stimulus. The test sequence then followed, randomized into a different order from that employed in the second stage of testing.

Results

Figure 1 summarizes the results obtained from the native speakers of English, and the Japanese students how were superior and inferior students of spoken English. For convenience, the results obtained in the first stage of testing with isolated [da]-[ga] stimuli are not included in this preliminary report, as the various groups did not differ in their perception of these sounds, and as the main interest is in the contrasting effects of [al] and [ar].

The native speakers of English (Figure 1a) were 100% correct in identifying [al] and [ar], and showed the anticipated context effect of [l] vs. [r]. The Japanese speakers who were superior students of spoken English (Figure 1b) were 99% correct in identifying [l] and [r], which confirms previous indications (MacKain, Best, & Strange, 1981) that at least some native speakers of Japanese can master the [l]-[r] distinction. Like the native speakers of English, these subjects showed the contrasting effects of [l] and [r] on perception of [da] and [ga]. In contrast to the other two groups of subjects, those Japanese subjects who were inferior students of spoken English (Figure 1c) averaged only 58% correct identification of [l] and [r], which is not significantly better than chance. Nonetheless, they showed the contrasting effects of [l] and [r] on perception of [da] and [ga]. Analysis of variance reveals significant main effects of stimulus number, $F(6,138) = 905.79$, $p < .0001$, and context, $F(1,23) = 31.93$, $p < .00001$, and an interaction of these two variables, $F(6,128) = 130.19$, $p < .00001$, but not interaction between subject group and context. There was also a main effect of subject group, $F(2,23) = 9.58$, $p < .0$[illegible], and an interaction involving subject group with stimulus number, $F(12,138) = 2.19$, $p < .00001$, and a small three-way interaction, $F(12,138) = 2.19$, $p < .015$. Each of these reflects the slightly aberrant behavior of the superior students of English in labeling the endpoints of the continuum.

Discussion

Thus, all of the subjects perceived some difference between spoken [l] and [r], and adjusted their perception of a "following" phoneme accordingly, whether or not they could phonetically identify [l] and [r]. If it is accepted that the context effect of [l] and [r] is specific to speech perception (Mann & Liberman, 1983) and reflects listeners' sensitivity to the acoustic consequences of coarticulation, one implication of the ability of the inferior students of spoken English to be sensitive to the effects of [l] and [r] while unable to identify them as phonemes, is that perception of speech errors on at least two levels: articulatory and phonological. The articulatory level is

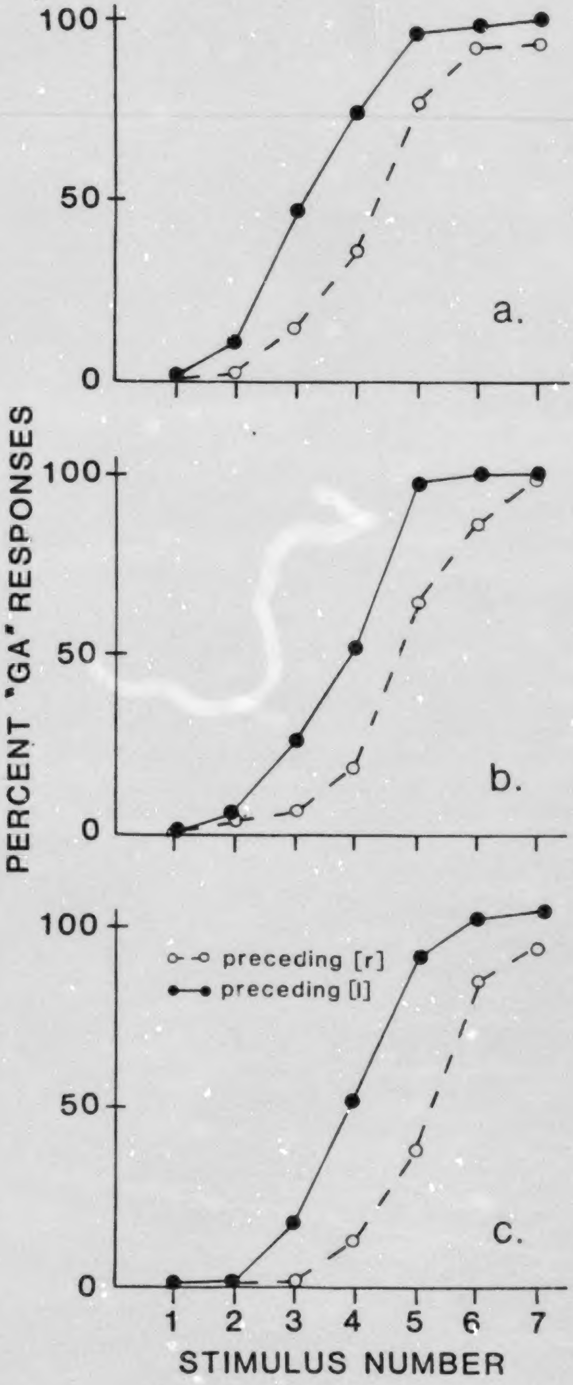


Figure 1. The contrasting effects of [l] and [r] on perception of the [d]-[g] distinction by: a) native speakers of English who are 100% correct in identifying [l] and [r]; b) native speakers of Japanese who are 99% correct in labeling [l] and [r], and c) native speakers of Japanese who perform at chance level in labeling [l] and [r].

directly responsible for those context effects and trading relations in speech perception that rest on the integration, interpretation, and abstract representation of incoming sensation as the product of human vocalization. The ability to represent speech sounds at this level is independent of native language experience; hence speakers are sensitive to the articulatory properties of the liquids [l] and [r] whether or not those phonemes are part of their native inventory. Moreover, articulatory representation may precede phonetic representation, as listeners may perceive articulatory differences that they cannot phonetically represent. As for the phonological level of representation, this higher level of speech perception may permit the phonetic identification of speech stimuli as [al] or [ar], [da] or [ga], and is available to consciousness. Unlike the articulatory level, however, it depends upon language experience; hence listeners may encounter difficulty when they are required to categorize consonants phonetically that are not in their native inventory.

In most speech perception experiments, subjects' responses are guided by the phonological level of representation. (Responses could also be mediated by a higher, lexical level of representation; for a discussion, see Forster, 1979.) Nonetheless, their behavior in identification and discrimination experiments has led to the view that speech is perceived as if by reference to the articulatory gestures that convey phonetic segments. Apparently the representation of those articulatory gestures occurs at a prior level that is less readily available to introspection (although it might become available through training). Were it to intervene directly upon consciousness, all Japanese subjects would be able to draw on their ability to perceive articulatory differences between [l] and [r], and thus be capable of distinguishing [l] and [r].

The distinction between phonological and articulatory representation of speech accounts for the ability and the inability of Japanese subjects to perceive a difference between [l] and [r], while reinforcing and extending some other observations in the speech literature. It is consistent with the fact that the context effect of [al] and [ar] is evident not only when subjects were required to label these utterances phonetically (Mann, 1980), but also when subjects are instructed to ignore them, as was the case for the native speakers of English in the present experiment. It also accords with evidence that subjects are sensitive to the articulatory properties of vowels that are not part of their native language (Whalen, 1981). Finally, it can offer a perspective on the interpretation of findings about the speech perception capabilities of infants. Infants have given evidence of perceiving many phonetically-relevant properties of utterances (see, for a review, Eilers, 1980; see also Kuhl, 1980; and Kuhl & Meltzoff, 1982), as well as evidence of trading relations (Miller & Eimas, 1983). It is clear that infants perceive human speech in a special way, perhaps owing to proclivities of the left or dominant hemisphere (MacKain, Studdert-Kennedy, Spieker, & Stern, 1983), which mediates speech perception in adults (Studdert-Kennedy & Shankweiler, 1970). At present, in the absence of any means of verifying that infants perceive phonemes, as such, it is premature to accept a conclusion that they are capable of phonological representation. Yet the data surely imply that infants possess some perceptual abilities that are the basis of adult phonetic perception (Miller & Eimas, 1983). One of these could well be the ability to form articulatory representations of incoming speech stimuli, regardless of specific language experience.

References

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, *29*, 191-211.

Eilers, R. E. (1980). Infant speech perception: History and mystery. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. II, pp. 23-39). New York: Academic Press.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, *18*, 341-347.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing*. Hillsdale, NJ: Erlbaum.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, *9*, 317-323.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. II, pp. 41-66). New York: Academic Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*, 1138-1144.

Liberman, A. M. (1982). On finding that speech is special. *American Psychologist*, *37*, 148-167.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, *30*, 133-143.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, *2*, 369-390.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. *Science*, *219*, 1347-1349.

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics*, *28*, 407-412.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, *14*, 211-235.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, *69*, 548-558.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, *13*, 135-166.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331-340.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, *9*, 283-303.

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, *55*, 678-680.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*, 81-110.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 621-637.

Studdert-Kennedy, M., & Shankweiler, D. P. (1970). Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, *48*, 579-594.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America*, *69*, 275-282.

A QUALITATIVE DYNAMIC ANALYSIS OF REITERANT SPEECH PRODUCTION: PHASE PORTRAITS, KINEMATICS, AND DYNAMIC MODELING*

J. A. S. Kelso,† Eric Vatikiotis-Bateson,†† Elliot L. Saltzman, and Bruce Kay†††

Abstract. The departure point of the present paper is our effort to characterize and understand the spatiotemporal structure of articulatory patterns in speech. To do so, we removed segmental variation as much as possible while retaining the spoken act's stress and prosodic structure. Subjects produced two sentences from the "Rainbow Passage" using reiterant speech in which normal syllables were replaced by /ba/ or /ma/. This task was performed at two self-selected rates, conversational and fast. Infrared LEDs were placed on the jaw and lips and monitored using a modified SELSPOT optical tracking system. As expected, when pauses marking major syntactic boundaries were removed, a high degree of rhythmicity within rate was observed, characterized by well-defined periodicities and small coefficients of variation. When articulatory gestures were examined geometrically on the phase plane, the trajectories revealed a scaling relation between a gesture's peak velocity and displacement. Further quantitative analysis of articulator movement as a function of stress and speaking rate was indicative of a language-modulated dynamical system with linear stiffness and equilibrium (or rest) position as key control parameters. Preliminary modeling was consonant with this dynamical perspective which, importantly, does not require that time *per se* be a controlled variable.

It has often been supposed that temporal organization in biological systems is ultimately governed by neural rhythm generators, biological clocks, metronomes, etc. Physiologists and psychologists, confronted with order in

---

*Journal of the Acoustical Society of America, in press.

†Also Departments of Psychology and Biobehavioral Sciences, The University of Connecticut.

††Also Department of Linguistics, Indiana University.

†††Also Department of Psychology, The University of Connecticut.

Acknowledgment. This research was supported by NIH Grant NS-13617, Biomedical Research Support Grant RR-05596, and Contract No. N00014-83-C-0083 from the U.S. Office of Naval Research. Part of the work was reported by J. A. S. Kelso and E. V.-Bateson at the 105th Meeting of the Acoustical Society of America in Cincinnati, OH in a paper entitled "On the cyclical basis of speech production" (*Journal of the Acoustical Society of America*, 1983, *73*, S67). We thank Betty Tuller and David Ostry for helpful comments on an earlier version of the paper and J. S. Perkell and W. L. Nelson for detailed and constructive reviews.

the time domain, have not hesitated to posit clocks whose "ticks" define when muscles will activate (e.g., Kozhevnikov & Chistovich, 1965; Rosenbaum & Patashnik, 1980). Our approach, however, has been directed towards identifying and understanding spatiotemporal pattern in articulatory events as a dynamic property of natural systems rather than as the result of the operation of some special neural or mental time-keeping device (cf. Kelso, Holt, Rubin, & Kugler, 1981). Once elaborated, we believe this dynamical perspective may afford a principled account of the ubiquity of temporal constraints in movement in general and in speech in particular. For example, the internal phasing relations among muscles and kinematic components in rhythmic activities such as locomotion, scratching, respiration, and mastication are preserved across scalar changes in force and rate (cf. Kelso, 1981; Grillner, 1982, for reviews). Similarly, in electromyographic and kinematic work on speech (Tuller, Kelso, & Harris, 1982, 1983; Tuller & Kelso, 1984), timing of consonant production relative to vowel production was found to be invariant over substantial changes (induced by stress and rate) in the duration of the vocalic cycle. These data--along with other evidence (reviewed by Fowler, 1983)--suggest a vowel-to-vowel organization that places constraints on speech timing.

Although speech certainly involves many of the same body parts as chewing, its rhythmic basis is not clear, in spite of the fact that linguists and others have long claimed speech to be rhythmic, and people perceive it to be so (e.g., Lehiste, 1972; Lenneberg, 1967; Lisker, 1975; Pike, 1945). Yet experimenters have had enormous difficulty identifying rhythmicity in either the articulatory or the acoustic domain. One possible reason--as pointed out by Fowler (1983) with respect to acoustic studies--is that experimental measurements typically used may be inappropriate for capturing the natural, temporal structure of spoken sequences. Speaking is an inherently multidimensional process; during speech different articulators are involved to different degrees and the spatiotemporal overlap among movements is considerable. Confronted with so many simultaneous or nearly simultaneous events, there seems little chance of our identifying any basic temporal regularity, even though our perceptual impressions lead us to suppose that one exists.

Our approach in the present work was to strip away, as much as possible, the influence of segmental variation on articulatory movement, by asking subjects to speak "reiterantly." That is, speakers substituted the syllable /ba/ or /ma/ for each real syllable in the utterance, while mimicking the utterance's normal prosodic structure. The benefit of the reiterant technique is that, by minimizing segmental variability while preserving the prosodic pattern (Liberman & Streeter, 1978; Nakatani, 1977), we are able to measure the movements of articulators (in this case the lips and jaw) that are consistently involved in the production of /ba/ and /ma/. In principle, this procedure affords an analysis of articulator patterns in a simple and accessible form.

We recognize that the relationship between real speech and reiterant speech is not always transparent. We should stress, however, that the main thrust of the present work is to use reiterant speech as a tool to examine articulator motions in a speechlike task. We do not claim any necessary generalization to real speech although one might exist (see also Larkey, 1983). For instance, Liberman and Streeter (1978) show the pattern of acoustic syllable durations to be similar between real speech and skilled reiterant speech although the absolute durational values are very different. In terms of production, it seems unlikely to us that the control of the lip-jaw system for the production of a reiterant /ba/ is fundamentally different when the same syllable is produced during natural speech. Indeed, we shall describe quantitatively certain kinematic relationships (e.g., between an articulator's peak velocity and displacement) that have been observed in many other nonreiterant speech production studies.

In the present paper, we outline a geometric approach for characterizing the dynamic properties underlying articulatory movements during reiterant speech. We use the phase portrait to facilitate the analysis of relevant articulatory variables when speakers produce these simple sequences of syllables. To our knowledge, phase portrait techniques have rarely been employed in speech production studies, even though their role is to describe the *forms of motion* in complex, multidegree-of-freedom systems (cf. Abraham & Shaw, 1982). Were one to count the neurons, muscles, and joints that cooperate to produce even a simple utterance, literally thousands of such elements would be involved. Yet normal speech is usually coherent and organized: A low dimensional pattern emerges from a system of high dimensionality that can be controlled with relatively few dynamic parameters.[1] Thus our approach is one in which we attempt to characterize regularities of articulator pattern in terms of a relatively abstract *functional* organization (cf. Kelso & Tuller, 1984a). We do not attempt to model peripheral biomechanics or neurophysiological mechanisms. Rather we use the phase portrait as a way of uncovering qualitatively the system's control structure and as a preface to a quantitative treatment of articulatory trajectories. In doing so we observe both invariant and systematically varying features of motion when stress and speaking rate are changed. Perhaps most important, our results, analyzed geometrically and interpreted from a dynamic perspective, do not require the assumption that time itself is a controlled variable. Instead, the form of articulator trajectories over time is seen as a consequence of a control structure whose dynamic parameters are functionally equivalent to those of a mechanical mass-spring system, namely: equilibrium (or rest) position, which is the position at which the net force on the mass is zero; and linear stiffness, which is the reactive force per unit displacement.

I. Methods and Procedures

Two adult speakers (one male [SK, the first author and a native speaker of an Ulster dialect of English], and one female [DW, a speaker of a New Jersey dialect of American English]) recited the first and last sentences of the "Rainbow Passage": (1) "When the sunlight strikes raindrops in the air, they act like a prism and form a rainbow," and (2) "There is, according to legend, a boiling pot of gold at one end." After reciting each sentence, speakers mimicked the prosodic pattern 2-4 times, substituting only /ba/ or only /ma/ for each syllable. So, for example, "When the sunlight strikes raindrops in the air" would be mimicked as "*ba* ba *ba* ba *ba* *ba* ba ba ba *ba*" (where underlining indicates a hypothetical stress pattern for the syllables). Upon completion of the task at a normal, conversational rate, it was then repeated at a faster rate. One of the speakers (SK) repeated this procedure at a later date. In all, 392 syllables at each rate were analyzed. We also obtained measures of each speaker's preferred frequency of jaw movement over an extended period of time, by asking the subject to "wag" the jaw at a comfortable amplitude and frequency "as if you were going to do it all day." "Wagging" movements were then sampled over a 30-s interval.

For speech and nonspeech tasks, vertical displacements of the lips and jaw were tracked using a device similar in principle to the commercially available SELSPOT system, which employs infrared LEDs that can be placed midsagittally on the nose, lips, and point of the chin. Modulated light from the diodes is captured by a camera equipped with a Schottky planar diode located in its focal plane. The output of the photodiode is fed to associated electronics that decode the signals and compute pairs of x and y coordinates. Up to eight channels of coordinate potentials may be generated simultaneously, each with a bandwidth of 0-500 Hz. These potentials are then fed to first-stage DC offset preamplifiers, which center the signals about the zero DC level. Following the offset adjustment, the coordinate values are transmitted via DC coupled amplifiers, checked by means of a monitoring oscilloscope, and recorded. Once the subject was seated with the LEDs in place, calibration was achieved by raising the camera a known distance (2 cm) and recording the output of the lower lip LED. Simultaneous acoustic recordings were also made. The movement data were recorded on FM tape and sampled at 200 Hz in later computer analysis. This included numerical smoothing (using a 25-ms triangular window), and differentiation (using a two-point central difference algorithm; James, Smith, & Wolford, 1977) for obtaining the derivatives of motion (velocity, acceleration).

Figure 1 shows an example of the position and velocity of the lower lip and jaw (i.e., the LEDs attached to lower lip and jaw) for the first part of sentence 1, "When the sunlight strikes raindrops in the air," where /ba/ is the reiterated syllable. In the movement traces, peaks and valleys denote the high and low vertical positions achieved by the indicated articulators. Thus peaks occur during lip closure for the bilabial stop and valleys occur during production of the low vowel /a/. In the velocity traces, peaks and valleys are the maximum velocities attained going into and out of a closure, respectively. The peaks and valleys were determined by a computer program which also calculated means (M) and standard deviations (SD) for peak-to-peak cycle duration and displacement and duration of opening (peak-to-valley) and closing (valley-to-peak) gestures.

II. Results and Discussion

Each of the following sections is designed to be self-contained in that a discussion accompanies each set of empirical findings. First we present data pertaining to the global temporal regularity of articulator movement that was observed in the experiments. Second, a qualitative dynamic analysis of articulatory motion is presented using the phase portrait to describe the forms of motion that are produced. Following is a quantitative kinematic analysis of motion and its derivatives that details effects of the local changes induced by stress and speaking rate transformations. We try to maintain continuity of presentation in this quantitative section by proceeding from lower-order to higher-order kinematic relations. Finally we present some of our preliminary efforts to model the present articulatory findings using an approach based in dynamical systems theory and supported by recent results in the field of physiological motor control.

A. Global Temporal Regularity

First we show separately for the two rates and two reiterant syllables the mean duration between successive peaks and the associated standard deviations. The values shown in Table 1 are averaged across subjects and sentences

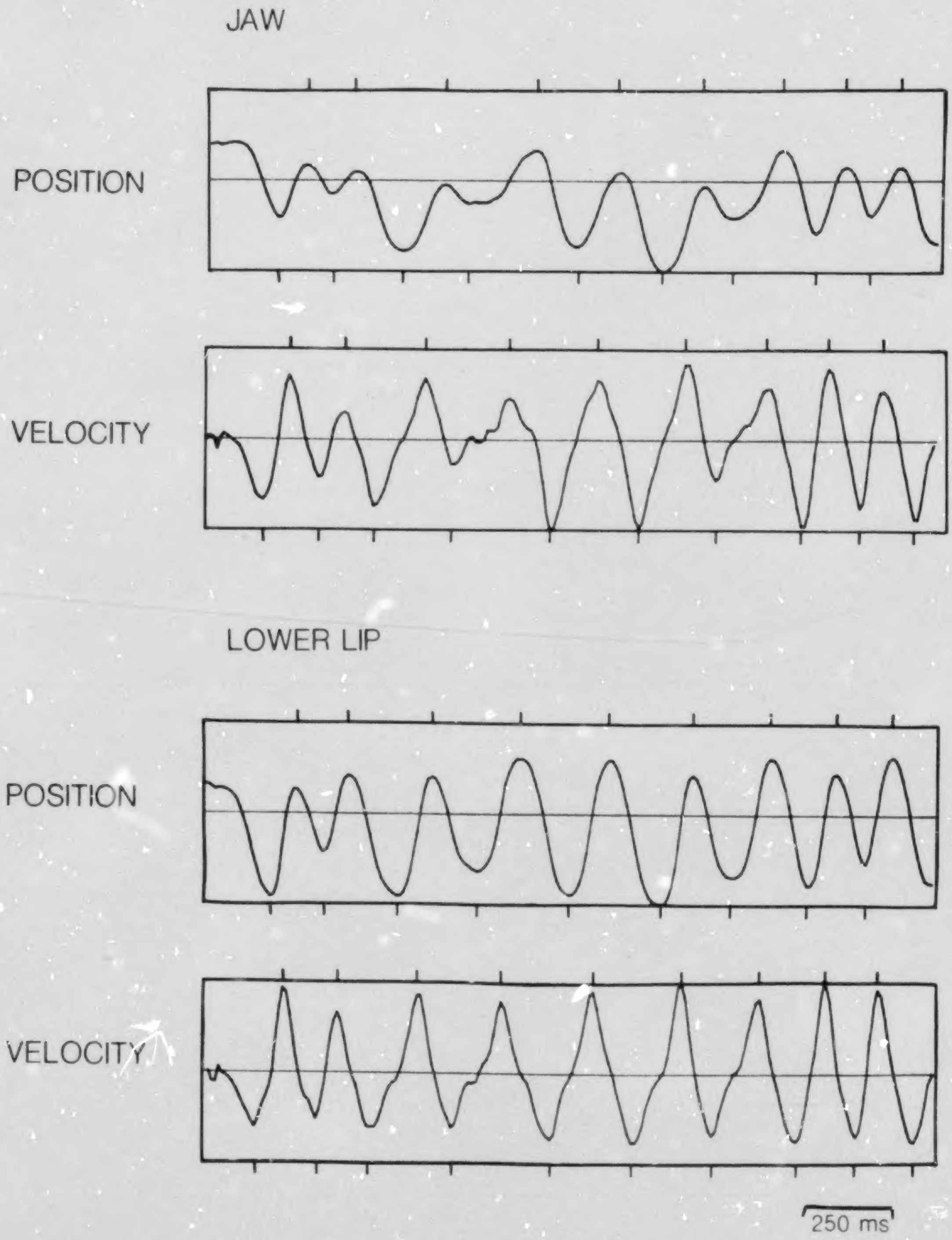


Figure 1. Position and velocity over time of lower lip and jaw LEDs for the reiterant production of "When the sunlight strikes raindrops in the air." /ba/ is the reiterant syllable.

for both jaw and lower lip motions (i.e., motions of the jaw and lower lip LEDs). In order to study articulatory motions per se, we have removed intervals that span major syntactic breaks and the first and last syllables of the sentence, i.e., where startup, pauses, and lengthening effects predominate.

Table 1

Means and standard deviations of peak-to-peak duration in ms and frequency (f) in Hz for jaw (J) and lower lip (LL) during reiterant speech at two rates. Between-subject standard deviations are in parentheses.

	/ba/		/ma/		OVERALL	
	J	LL	J	LL	J	LL
NORMAL						
m	213 (5)	212 (8)	212 (3)	211 (3)	212 (4)	211 (6)
sd	42 (6)	40 (6)	41 (1)	37 (1)	42 (4)	39 (5)
f	4.70 (.11)	4.72 (.19)	4.72 (.06)	4.73 (.06)	4.72 (.08)	4.73 (.14)
FAST						
m	168 (5)	168 (5)	166 (3)	165 (4)	167 (4)	167 (4)
sd	33 (9)	29 (8)	29 (3)	30 (5)	31 (7)	30 (6)
f	5.95 (.17)	5.95 (.18)	6.03 (.11)	6.06 (.15)	5.98 (.14)	6.00 (.16)
n	512	512	272	272	784	784

The durational data show quite low variability regardless of rate, with coefficients of variation in the 10% to 20% range. The two speakers are also very similar in their durational behavior as revealed in the small between-subject standard deviation of the means. Mean cycle durations for the three experimental sessions were 211 ms (approximately 5 Hz) for the normal rate and 167 ms (approximately 6 Hz) for the fast rate. In this case, the jaw exhibits a periodicity similar to that of the lower lip. Not surprisingly, the data contrast with those of Ohala's (1975) earlier study in which 10,000 consecutive jaw opening gestures were obtained during a 1.5-h reading period. Ohala (1975) found large durational variance (presumably because of the presence of pauses and segmental factors) accompanied by a dominant, but weakly defined, periodicity of about 250 ms (4 Hz). Ohala and others (e.g., Lindblom, 1983) have suggested that this periodicity may correspond to the "preferred frequency of the mandible." However, the preferred wagging frequencies of jaw movement for our two speakers (0.81 Hz and 2.04 Hz, SDs = 0.06 and 0.21 Hz, respectively) are much slower than the frequencies found either by us for reiterant speech or by Ohala for read speech. It is clear then, that neither the sharply defined periodicity observed by us in reiterant speech nor the weakly defined cycling found by Ohala in read speech is the same as the preferred frequency of the mandible in our nonspeech task (see also Nelson, Perkell, & Westbury, 1984, for differences between preferred frequencies of mandible movement in speechlike and nonspeech tasks). We also found that the periodicity was unaffected by the syllable that was used to mimic real speech. The largest mean durational difference regardless of rate condition between /ba/ and /ma/ for any articulator was 3 ms (see Table I). In short, when segmental variation is minimized, it is possible to identify a relatively stable articulatory periodicity. The periodicity is not *perfectly* isochronous because there are systematic variations concomitant with stress and rate (see Section IIC).

B. A Geometric (Qualitative Dynamic) Analysis[2]

In the following geometric analysis, phase plane trajectories are generated by continuously plotting the relationship between, in this case, articulator position, $\underline{x}$, and its derivative, velocity, $\underline{\dot{x}}$. As an example, consider the idealized case shown in Figure 2. The upper trace is a computer generated sinewave of 5 Hz with a peak-to-valley displacement defined to be 20 mm. The peak position corresponds to the consonant closure, and the valley position to the maximum opening for the vowel. Points of maximum downward (opening) and upward (closing) velocity fall at the midpoints of the position trace. To create a phase plane trajectory shown on the lower part of Figure 2, we plot successive position points and their corresponding velocities as coordinates on a plane whose vertical axis denotes position and whose horizontal axis denotes velocity.[3] The arrowheads on the circle denote the direction of motion on the plane. Thus one cycle or orbit corresponds to the interval between successive closures, with the opening gesture on the left half and the closing gesture on the right. *Note that time itself is not an explicit variable in this description.*

Figure 3 shows phase plane trajectories for the jaw and lower lip LEDs of "When the sunlight strikes raindrops in the air," using reiterant /ba/ spoken at a normal rate. Qualitatively, the shapes of the trajectories are quite similar across the ten syllables plotted. There is a strong tendency, for example, for displacement and peak velocity to covary directly (see Section IIC). Normal and fast reiterant productions for subjects SK and DW of the second part of the first sentence, "they act like a prism and form a rainbow," are shown in Figures 4 and 5. The mutual relationship between the kinematic variables of position and velocity is accentuated by the rate manipulation, particularly for subject SK. Once again, even when there is a clear distinction between the trajectories corresponding to stressed and unstressed syllables, their orbital shapes are generally similar. The unstressed (sometimes reduced) syllables are characterized by smaller displacements and peak velocities than the stressed syllables, thus maintaining a global similarity of (elliptical) trajectory shape across unstressed and stressed gestures. Also observed, however, are subtle differences between trajectory shapes associated with different gestural displacements. For example, the orbits appear to be slightly more compressed horizontally for larger displacement gestures relative to shorter displacement gestures. In Section IIC, we will quantify both the global similarities and subtle differences among gestural trajectory shapes.

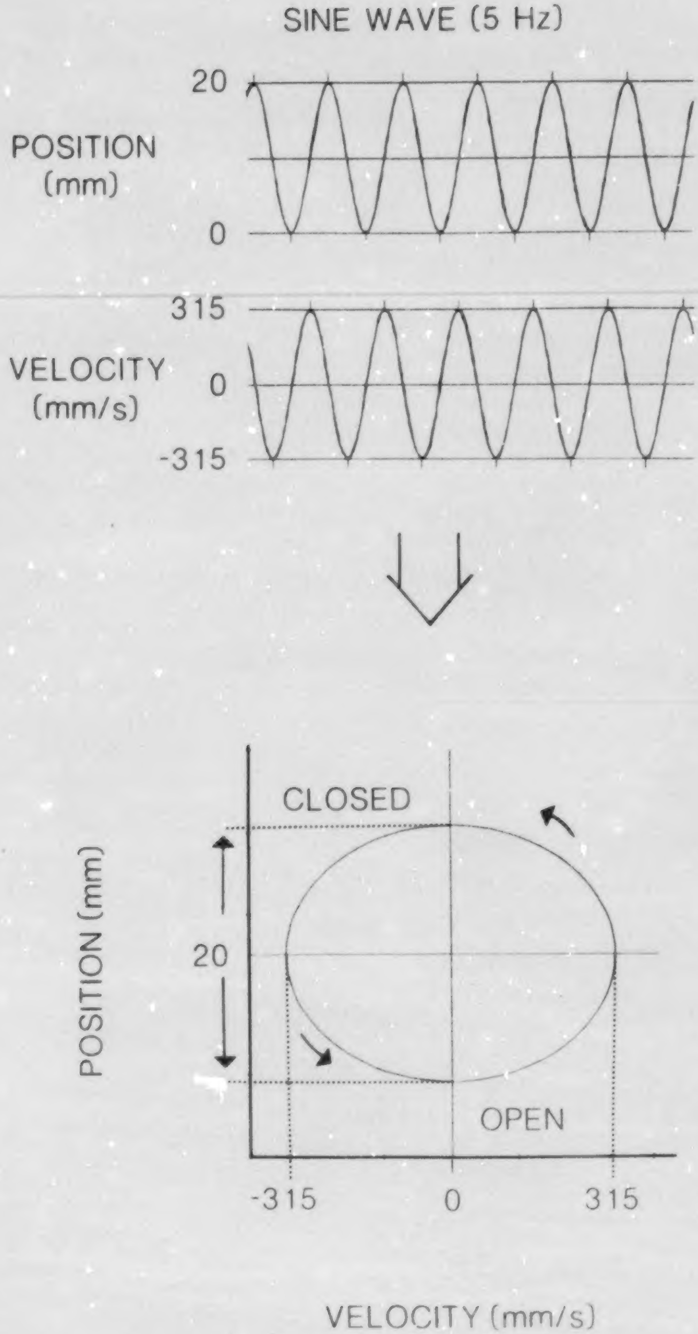


Figure 2. Top. Idealized position and velocity over time of articulator movement. Bottom. Corresponding phase plane trajectories. Abscissa is velocity, ordinate is position (see text for details).

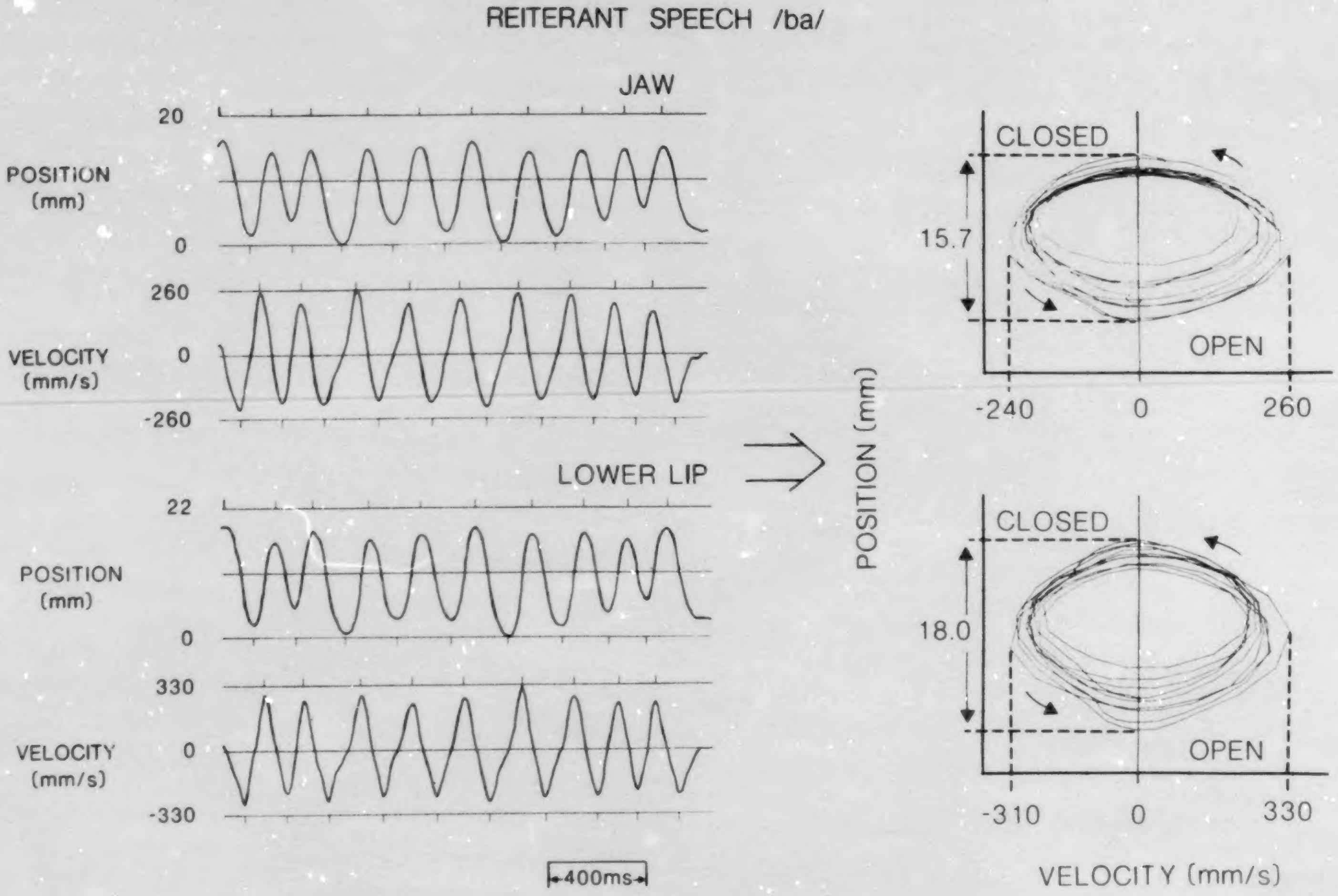


Figure 3. Left. Position and velocity over time of jaw and lower lip LEDs for sentence produced with reiterant /ba/ at a normal rate. Right. Corresponding phase plane trajectories.

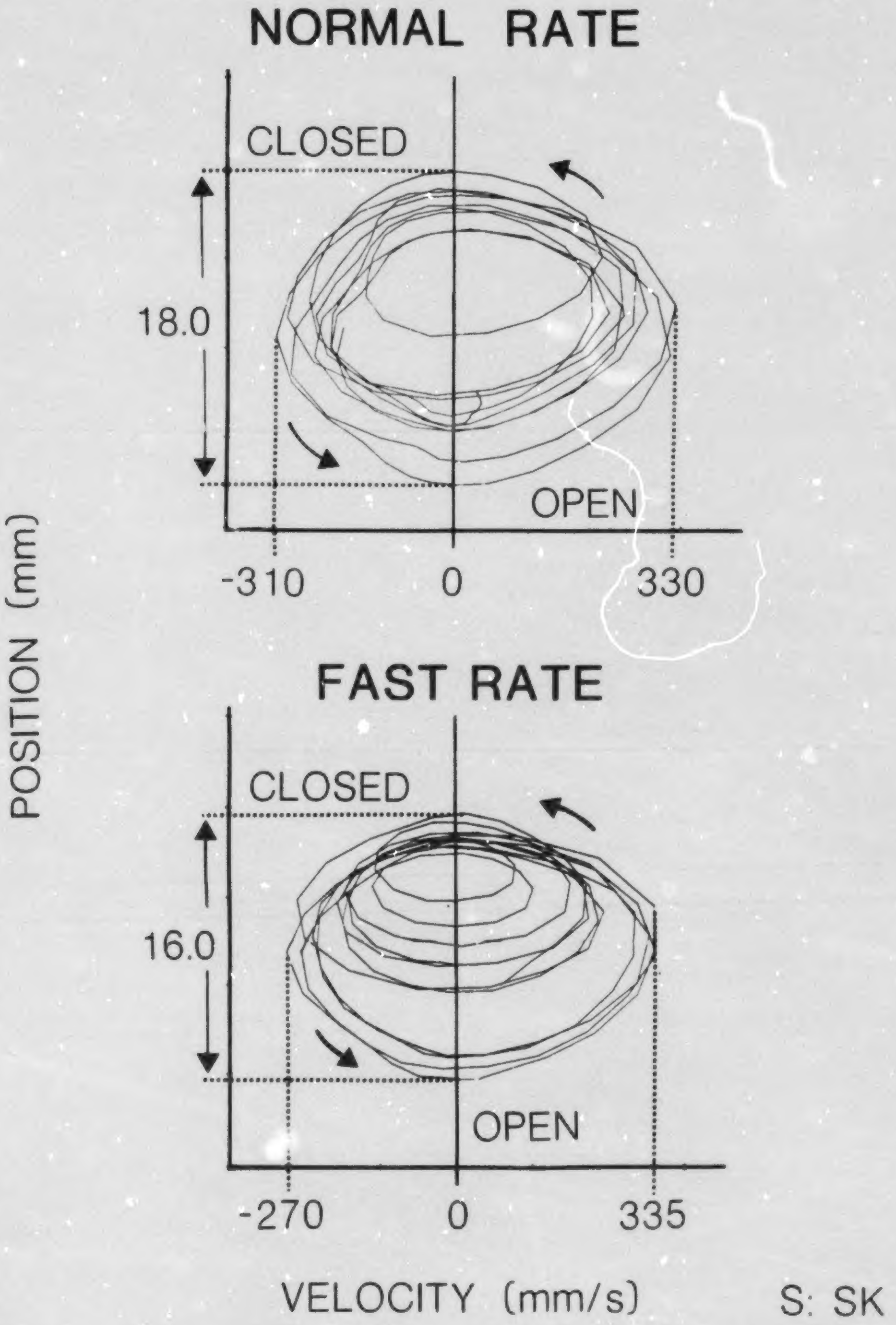


Figure 4. Phase plane trajectories of lower lip motions for the second part of sentence 1, "They act like a prism and form a rainbow" produced at normal and fast speaking rates with /ba/ as the reiterant syllable. Subject is SK.

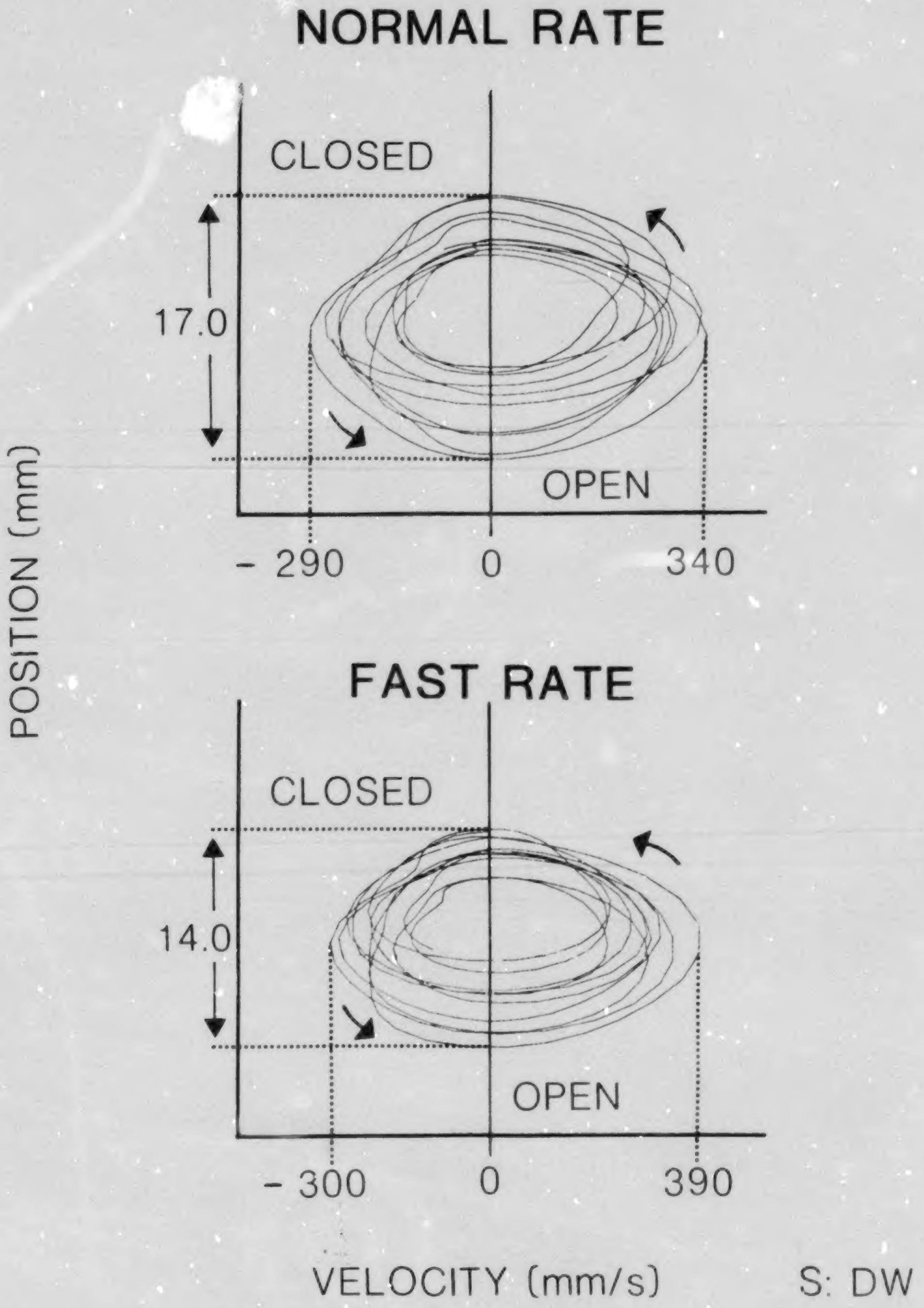


Figure 5. Phase plane trajectories of lower lip motions for the second part of sentence 1, "They act like a prism and form a rainbow" produced at normal and fast speaking rates with /ba/ as the reiterant syllable. Subject is DW.

C. Quantitative Kinematic Analysis

In this section we quantify specific effects of speaking rate and stress on articulatory movements in an effort to answer the following questions. First, what kinematic variables or relations among variables might inform us about the control of speech gestures? Second, what kind of regularity, if any, exists in the motions of speech articulators across changes in stress and speaking rate, and how might such regularity be rationalized? Although we appreciate that there are many idiosyncratic differences among speakers, dialects, and languages, our emphasis here is on identifying what is common across such diversity. In short, can we begin to define a "deep structure" for speech motor control that can be recognized in the face of much surface variability, and, if so, on what principle(s) is it based?

We begin with an analysis of the space-time characteristics of articulator movement and its derivatives, with the emphasis now on the gesture (opening and closing) rather than the cycle. Because of the enormous amount of kinematic data involved, we restrict our concerns (unless otherwise indicated) to (a) the motions of the jaw and lower lip complex for the syllable /ba/ during reiterant speech, and (b) the single experimental session for each speaker, i.e., omitting the repeated session. This amounts to 232 gestures for speaker DW (116 opening and 116 closing) and 464 gestures for speaker SK (232 opening and 232 closing).

The general statistical analysis of the kinematic variables takes the form of a gesture (opening, closing) X stress (stressed, unstressed) X rate (normal, fast) analysis of variance for each dependent variable, followed by correlational analysis between variables (e.g., displacement versus time) where appropriate. In order to facilitate communication of the results we report the degrees of freedom for the statistical main effects and interactions only once. For subject DW the numerator and denominator degrees of freedom are 1 and 224; for subject SK they are 1 and 456.

1. Displacement, movement time, and their relation.

Tables 2 and 3 provide the mean displacement and mean movement times of the opening and closing gestures for the syllable /ba/, as a function of speaking rate and stress. The mean data order systematically for both kinematic variables in both subjects, although the magnitude of change across rate and stress is idiosyncratic. Similar results have been reported by others (e.g., Kuehn & Moll, 1976; Tuller, Harris, & Kelso, 1982b). For displacement, since the lips always return to closure, the main effect of gesture type (opening versus closing) was not significant in either subject's data; $Fs = 0.10$ and 0.55, $ps > 0.05$ for DW and SK, respectively. Nor were there two- or three-way interactions with gesture type. Stressed gestures had larger displacements than unstressed gestures; $Fs = 39.19$ (DW) and 415.44 (SK), $ps < 0.0001$. Rate had a generally similar effect: Normal rate gestures were produced with larger displacements than fast gestures, $Fs = 11.26$ (DW) and 136.18 (SK), $ps < 0.001$. Unlike DW, subject SK revealed a stress X rate interaction on the displacement measure, $F = 35.44$, $p < 0.0001$. A simple main effects analysis of this interaction was entirely consonant with the main effects, however: The difference in displacement as a function of rate was more apparent in unstressed gestures, $F = 162.92$, $p < 0.0001$, than stressed gestures, $F = 8.70$, $p < 0.004$. Similarly, differences in displacement as a function of stress were manifest particularly at a fast speaking rate, $F = 346.77$,

Table 2

Kinematic values of displacement, time, and peak velocity across rate and stress variations (opening gestures, /ba/)

			Stressed			Unstressed		
			d[1]	t	Vp	d	t	Vp
Normal	DW	M	14.58	123.9	229.2	11.80	112.4	204.0
		SD	3.68	20.6	74.2	3.30	22.4	65.9
		n	24			34		
	SK	M	16.02	140.4	262.5	12.63	114.7	238.6
		SD	1.40	19.0	24.9	3.35	30.1	55.8
		n	48			68		
Fast	DW	M	13.41	103.3	241.0	10.38	85.3	216.3
		SD	2.83	12.7	48.8	3.27	11.3	63.6
		n	24			34		
	SK	M	14.85	120.0	241.1	8.27	81.3	170.2
		SD	1.46	17.9	32.9	3.85	20.6	64.4
		n	48			68		

[1] Displacement (d) in mm; Time (t) in ms; Peak Velocity (Vp) in mm/s

Table 3

Kinematic values of displacement, time, and peak velocity across rate and stress transformations (closing gestures, /ba/)

			Stressed			Unstressed		
			d[1]	t	Vp	d	t	Vp
Fast	DW	M	13.88	91.8	297.0	12.14	82.5	265.3
		SD	2.94	17.8	73.6	2.91	20.1	74.7
		n	24			34		
	SK	M	15.99	106.4	321.7	12.09	84.8	272.2
		SD	1.26	15.1	19.2	2.77	14.9	52.8
		n	48			68		
Fast	DW	M	13.07	66.3	348.4	10.33	64.7	280.5
		SD	2.39	14.8	65.6	3.05	11.1	96.5
		n	24			34		
	SK	M	14.88	86.5	323.0	8.17	66.5	202.3
		SD	1.41	12.1	27.7	3.11	10.6	68.1
		n	48			68		

[1] Displacement (d) in mm; Time (t) in ms; Peak Velocity (Vp) in mm/s

$p < 0.0001$, although they were highly significant at the normal rate as well, $F = 104.11$, $p < 0.0001$ (see Tables 2 and 3). No other interactions were significant for either subject.

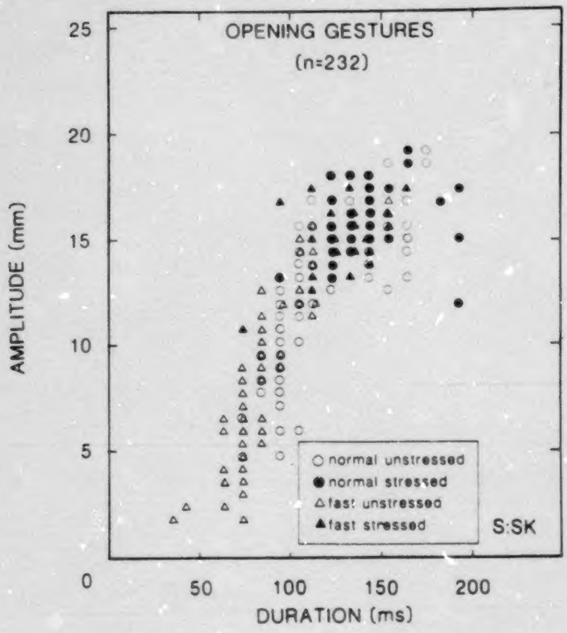
For movement time, opening gestures as a class took longer than closing gestures. All the movement time values for similar conditions reported in Table 2 (opening) are greater than those reported in Table 3 (closing), a finding substantiated by a significant gesture main effect for DW, $F = 171.43$, $p < 0.0001$ and SK, $F = 240.57$, $p < 0.0001$. Stressed gestures take longer than unstressed gestures, Fs = 20.21 and 223.62, ps < 0.0001 for DW and SK, respectively. For subject DW a simple main effects analysis of the significant gesture X stress interaction, $F = 4.31$, $p < 0.04$ revealed that the stress effect was greatest for opening gestures, $F = 21.58$, $p < 0.001$ (compare Tables 2 and 3). For subject SK, the gesture X stress interaction was also significant, $F = 10.34$, $p < 0.002$: The difference in movement time between stressed and unstressed conditions was greater for opening gestures, $F = 165.08$, $p < 0.0001$, than closing gestures, $F = 68.89$, $p < 0.0001$.

Speaking rate had a systematic effect on movement time. Gestures produced at a normal rate took longer than those at a faster rate, $F = 104.50$ (DW) and $F = 181.84$ (SK), ps < 0.0001. For subject SK, there was also a gesture X rate interaction, $F = 6.60$, $p < 0.02$. Again, the rate effect between gestures was a matter of degree; movement time differences between rates were more apparent in opening gestures, $F = 128.86$, $p < 0.001$ than closing gestures, $F = 59.98$, $p < 0.0001$, although clearly the effect was highly significant in both gesture types.

In summary, in both subjects, the main effects of stress and rate predominate for both displacement and movement time as dependent measures, although these effects tend to be greater in opening gestures than closing gestures. Generally speaking, stressed gestures display greater articulatory displacement and longer duration than unstressed gestures. Rate has similar effects. Gestures produced at faster speaking rates are accomplished with smaller displacements and in shorter movement times than those at a normal conversational pace.

Viewed from an overall perspective based on the mean data of each subject, we can make a rather simple statement regarding the displacement-time relation independent of movement phase (opening versus closing), rate, or stress. Namely, on the average, *displacement covaries directly with duration*. Smaller (larger) displacements tend to be observed at fast (normal) rates and in unstressed (stressed) environments; duration of motion adjusts in a corresponding fashion.

These overall effects, therefore, suggest a systematic and apparently quite linear relationship between spatial and temporal dependent measures. However, examination of the scatter plots for each subject in Figure 6 (opening phase) and Figure 7 (closing phase) reveals a somewhat more complicated picture. For subject SK the data follow the general picture outlined above; amplitude and duration vary in a quite linear way. The overall correlation for opening gestures is $r = 0.82$ and for closing gestures $r = 0.76$ (ps < 0.01). Moreover, the displacement-time correlations for individual conditions, shown in Table 4, are, with a single exception, significant (ps < 0.05).



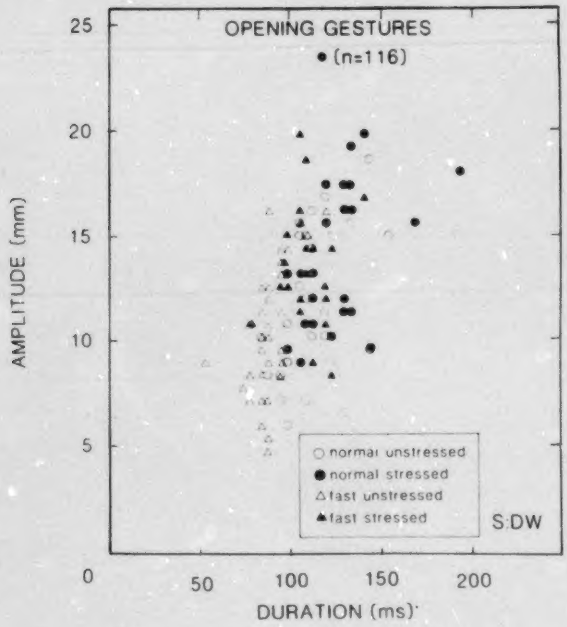


Figure 6. Scatter plot of amplitude and duration of each subject's lower lip motions for opening gestures associated with the consonant-vowel (CV) portion of the syllable. Points are differentiated by rate and stress, as shown in legend.

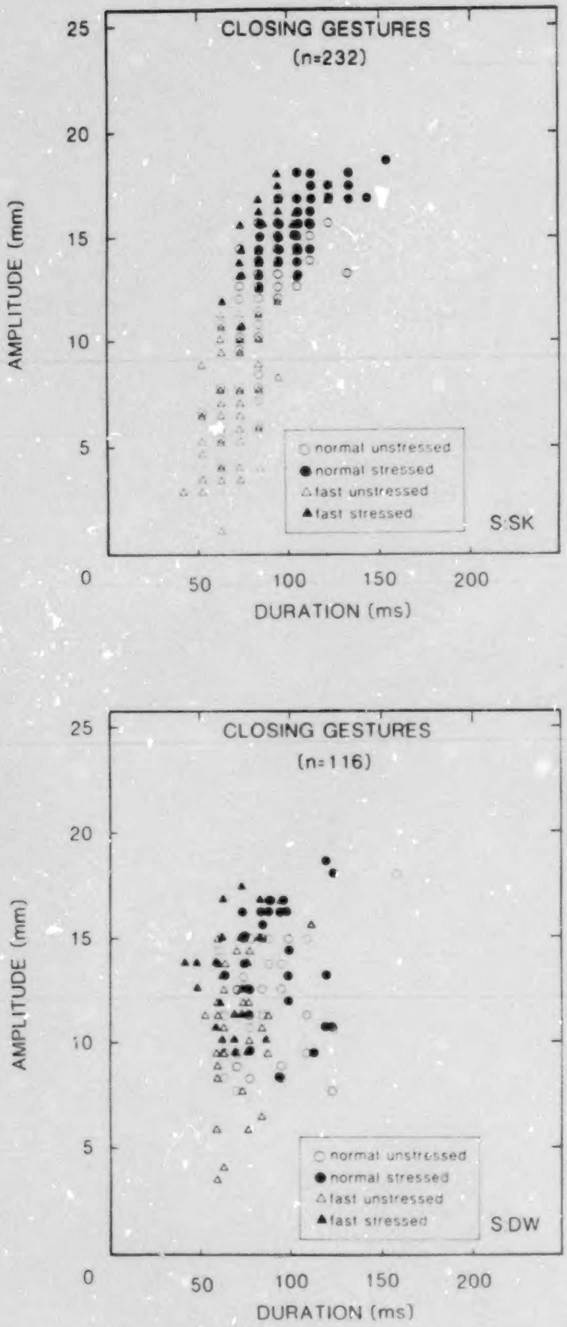


Figure 7. Scatter plot of amplitude and duration of each subject's lower lip motions for closing gestures associated with the vowel-consonant portion (VC) of the syllable. Points are differentiated by rate and stress, as shown in legend.

Table 4

Linear correlations (r) and regression slopes (m) of displacement-time relationship across rate and stress transformations (/ba/).

A. Opening Gestures

		Stressed		Unstressed	
		m	r	m	r
Normal	DW	.06	.33	.07	.46*
	SK	.02	.24	.08	.72*
Fast	DW	.01	.05	.10	.35
	SK	.05	.55*	.15	.82*

B. Closing Gestures

		Stressed		Unstressed	
		m	r	m	r
Normal	DW	.02	-.11	.05	.31
	SK	.05	.60*	.12	.65*
Fast	DW	.02	.23	.02	.08
	SK	.06	.50*	.15	.52*

*$p < .05$

The picture is rather different for subject DW, however. In her data the individual displacement-time pairs are widely distributed and in only one out of a possible eight conditions (unstressed opening gestures produced at a normal rate) is there a significant correlation (see Table 4). When opening and closing gestures are analyzed as a group for DW, significant correlations are obtained, rs = 0.46 and 0.26 (ps < 0.05), respectively, although the proportion of variance accounted for is small.

To summarize, the coupling between displacement and time is quite different for the two subjects. One subject (SK) reveals a rather orderly relation between these variables across rate, stress, and movement phase (opening vs. closing). The other subject (DW) shows a high degree of overlap among conditions and a much more homogeneous distribution of displacement-time data pairs. Indeed, the proportion of variance accounted for by this relationship is so small as to suggest that, for DW, displacement and time are essentially independent.

How might these apparent discrepancies between subjects in the displacement-time performance space be interpreted? One account that merits mention is that the speech motor system adheres to a minimum cost function such as "least effort," which might give rise to tradeoffs in articulatory displacement and duration. This notion of movement costs is elaborated in some detail in a recent paper by Nelson (1983) and has been applied to an analysis of jaw movements in repetitive speech and nonspeech gestures (Nelson et al., 1984). The key idea is that articulatory movements during speech are accomplishing system "goals" in the physically most economical fashion, i.e., according to some "ease of movement" criteria (see also Lindblom, 1983), which in turn imposes boundary constraints on speech motor programming (Nelson, 1983). Such criteria may be met by minimizing a number of possible articulatory cost indices such as "effort" (proportional to peak velocity, which bears a direct relation to the impulse or integral of the force-time curve for a given movement) or "jerk" (the first derivative of acceleration). Nelson (1983) shows that although a wide variety of "movement ease" cost functions may be minimized, the displacement-duration relation remains roughly the same. Thus a common feature of all such functions is that "cost" increases (on whatever dimension) are associated with moving a given distance in less time or moving a greater distance within a given time. To do either requires an increase in peak velocity, acceleration, jerk, etc. (see also Hogan, 1984).

In the displacement-time space a relationship, such as that displayed by subject SK in Figures 6 and 7 is suggestive of a fairly constant articulatory cost (cf. Nelson, 1983, Figure 5). Thus it could be argued that gestures of short amplitude and duration (e.g., fast unstressed gestures) do not necessarily cost the system any more than larger amplitude movements of greater duration (corresponding, say, to normal stressed gestures). Distance and time mutually adapt to the linguistic requirements of the activity in such a way as to preserve a relatively constant cost.

A problem, however, with this analysis of "economy of effort" in speech is that it appears to pertain, at best, to only one of our subjects and to only one of the three subjects in the Nelson et al. (1984) study. Several possibilities could account for such a state of affairs. One is that it could reflect differences in the skill level of producing reiterant speech. That is, the less constrained, more variable relation between displacement and time in subject DW suggests that her mode of motor control is not following a strategy of minimum cost. DW may, in fact, have to discover exactly what that strategy is. It is well appreciated in the literature (e.g., Larkey, 1983) that reiterant speech is itself a skill, and it was certainly our impression that subject DW was not as skilled at "converting" real speech into reiterant speech as was subject SK. How cost functions change with increasing skill is a topic open to much further research.

Given that the displacement-time relation is not consistent between subjects in the present study or in the literature in general (see Nelson et al., 1984; Parush, Ostry, & Munhall, 1983; Tuller et al., 1982b), the question is: Are there other observables that might afford insight into the similarity among subjects in this task? Are subjects really as different in performing reiterant speech as the displacement-time distributions suggest? As we shall see, examination of the higher derivatives of motion not only affords a window into the nature of the system's underlying dynamic organization, but also suggests that the differences between subjects might be due to the surface nature of the displacement-time description.

2. Peak velocity and the peak velocity-displacement relation

The phase plane data discussed in Section IIB reveal at least two interesting features about a given gesture's velocity pattern that merit further quantification. First, the patterns are largely unimodal (see Figures 3, 4, and 5) in that both opening and closing gestures possess single velocity peaks. Related to this, peak velocity (Vp) bears a direct relationship to total impulse (i.e., the integral of the force magnitude as a function of time), and thus can usefully be used to index the "effort" underlying the movement (e.g., Nelson, 1983; Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979). Since variables like stress have been associated with articulatory effort (e.g., Ohman's, 1967, stress pulse theory) it is of interest to quantitatively assess if and how peak velocity changes with gesture type, rate, and stress conditions. Second, and perhaps more important, is the apparent regularity--evident on the phase plane--in the covariation between a gesture's peak velocity (Vp) and its displacement (d). We consider first the statistical effects on peak velocity itself; then we evaluate and interpret the relationship between peak velocity and displacement.

A cursory look at Tables 2 and 3 indicates that Vp, like displacement and movement time, varies systematically with stress and rate, although in somewhat idiosyncratic ways. The gesture type main effect is significant for both subjects, $Fs = 59.08$ and 111.01, $ps < 0.0001$, for DW and SK, respectively. For similar conditions, all the Vp values in Table 3 (closing) are greater than Table 2 (opening). Stress had predictable effects on peak velocity regardless of gesture type. As in the recent results of Stone (1981) on jaw movement, and Ostry, Keller, and Parush (1983) on tongue dorsum movement, stressed gestures are produced with higher peak velocities than unstressed gestures, $F = 15.03$, $p < 0.002$ and $F = 201.48$, $p < 0.0001$, for DW and SK, respectively.

As others have found, however, the effect of speaking rate on the Vp measure was not so consistent across subjects (e.g., Abbs, 1973; Kuehn & Moll, 1976; Ostry et al., 1983, Tuller, Harris, & Kelso, 1982b). For subject DW, peak velocity was greater for the faster speaking rate, $F = 4.94$, $p < 0.03$. For SK, the opposite occurred (see Tables 2 and 3), $F = 94.41$, $p < 0.0001$. Moreover, there was a stress X rate interaction for SK, $F = 40.06$, $p < 0.0001$ but not DW, $F = 0.86$, $p > 0.05$. For SK, although stressed gestures are always produced more rapidly than unstressed gestures in both speaking rates, $F = 30.93$, $p < 0.0001$ (normal) and $F = 210.61$, $p < 0.0001$ (fast), only unstressed gestures differentiate between normal and fast speaking rates. For subject SK's unstressed gestures, normal speaking rates have higher Vp than fast speaking rates, $F = 132.50$, $p < .0001$. For stressed gestures, no significant differences in Vp occur between speaking rates (see Tables 2 and 3), $F = 1.97$, $p > 0.05$.

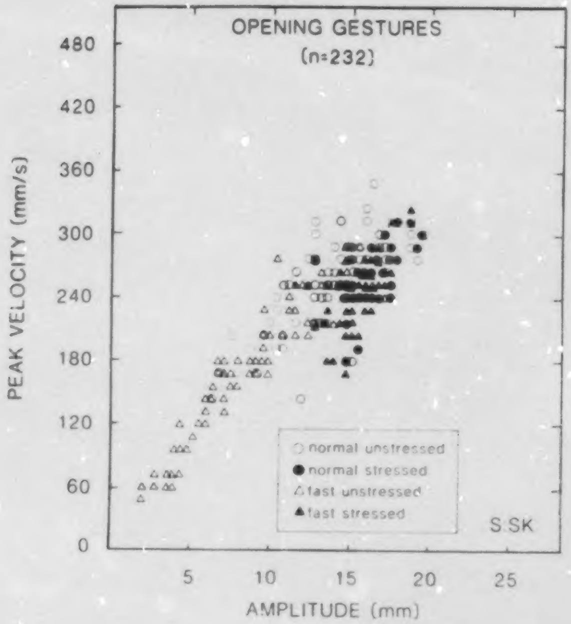
Because stress has very systematic effects on a variety of variables (including not only the kinematics reported here, but EMG as well, e.g., Tuller, Harris, & Kelso, 1982a) and the effects of rate are less systematic across subjects (particularly for Vp), it can be argued that stress and rate are qualitatively different kinds of articulatory transformations (see Tuller et al., 1982a, for review). However, the differences observed between stress and rate remain puzzling at least when viewed on single dimensions (e.g., EMG amplitude, duration, and articulator velocity), and further work is necessary to establish the validity of this claim. One potential issue--yet to be fully explored--is that the subject is usually free to vary the elected rate whereas stress constraints are more clearly defined. Systematic control of speaking rate may prove useful and enlightening.

The linkage between peak velocity and displacement, however, is less ambiguous. This finding in itself is not new; it has been reported before in other studies of articulation, often as an incidental result (e.g., Kent & Moll, 1972; Kozhevnikov & Chistovich, 1965; Kiritani, Imagawa, Takahashi, Masaki, & Shirai, 1982; Kuehn & Moll, 1976; Ohala, Hiki, Hubler, & Harshman, 1968; MacNeilage, 1970; Perkell, 1969; Sussman, MacNeilage, & Hanson, 1973). The particular articulator involved does not appear to be a factor; the relationship exists in movements at the supralaryngeal level including the tongue dorsum, tongue tip, lips, and jaw. More recently it has been observed in both abduction and adduction of the vocal folds (Munhall & Ostry, 1984).

Quantitative analysis of the present data reveals that Vp and d are highly correlated for opening and closing gestures in both subjects. For subject SK the correlations, collapsed across stress and rate, are 0.87 for the opening phase and 0.94 for the closing phase. For subject DW the correlations are 0.84 and 0.76 for opening and closing gestures, respectively ($\underline{p}s < 0.01$). Compared to the displacement-time relationship, which was very different between subjects (cf. Figures 6 and 7), the scatter plots displayed in Figures 8 and 9 for opening and closing gestures, respectively, show a much greater degree of overall similarity between subjects in both phases of motion.

The high linearity, of course, is a reflection of the overall temporal stability present in the opening and closing phases of the articulatory movements across rate and stress transformations. Since the slope of the Vp-d relationship for a given gesture type can be expressed in units of frequency, a perfect correlation between the two variables would indicate that the opening or closing gestures were of the same frequency, i.e., were perfectly isochronous. There are, however, local effects of stress and rate when the data are partitioned into subcategories, as can be seen from the absolute values of displacement, peak velocity, and duration given in Table 2 for opening gestures and Table 3 for closing gestures. In Table 5 we present the linear regression slopes and correlations of the peak velocity-displacement relationship for opening and closing gestures as a function of stress and rate. Overall, although the correlations are generally high and significantly different from zero, the slopes of the relationship between peak velocity and displacement are quite variable across subcategories. How might the slopes of the kinematic relation between Vp and d be interpreted with respect to the control processes underlying the reiterant speech task? First we address the significance of the overall Vp-d relation, then we consider the specific effects of rate and stress.

Recent theoretical considerations and empirical findings in the motor control field support an account of the Vp-d relation that is based on a movement's dynamics, not its kinematics. Relations among kinematic variables are useful to describe the space-time behavior of articulators, but it is dynamics that cause such motions. That is, it is important to realize that changes in displacement and its time derivatives (velocity and acceleration) are consequences of dynamical systems with parameters such as mass, stiffness, and damping. It is possible, however, to infer the structure of the underlying dynamics from the kinematics of articulator motions during either discrete or rhythmic tasks.



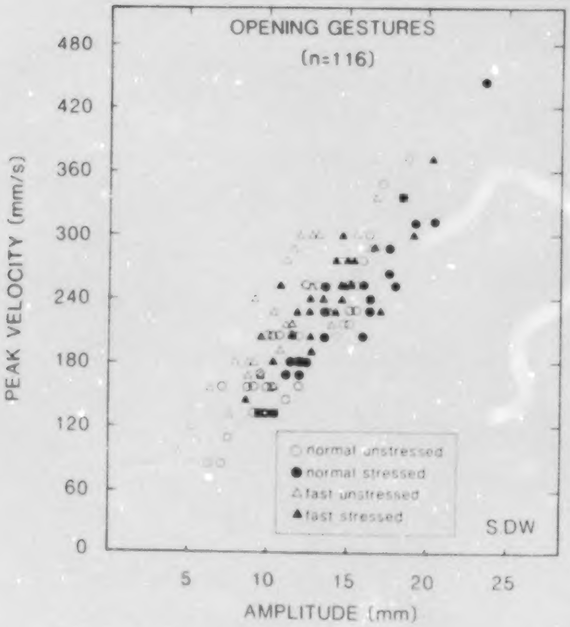


Figure 8. Scatter plot of peak velocity versus (peak-to-valley) amplitude (lower lip) of each subject's opening gestures associated with the consonant-vowel portion of the syllable. Legend specifies conditions.

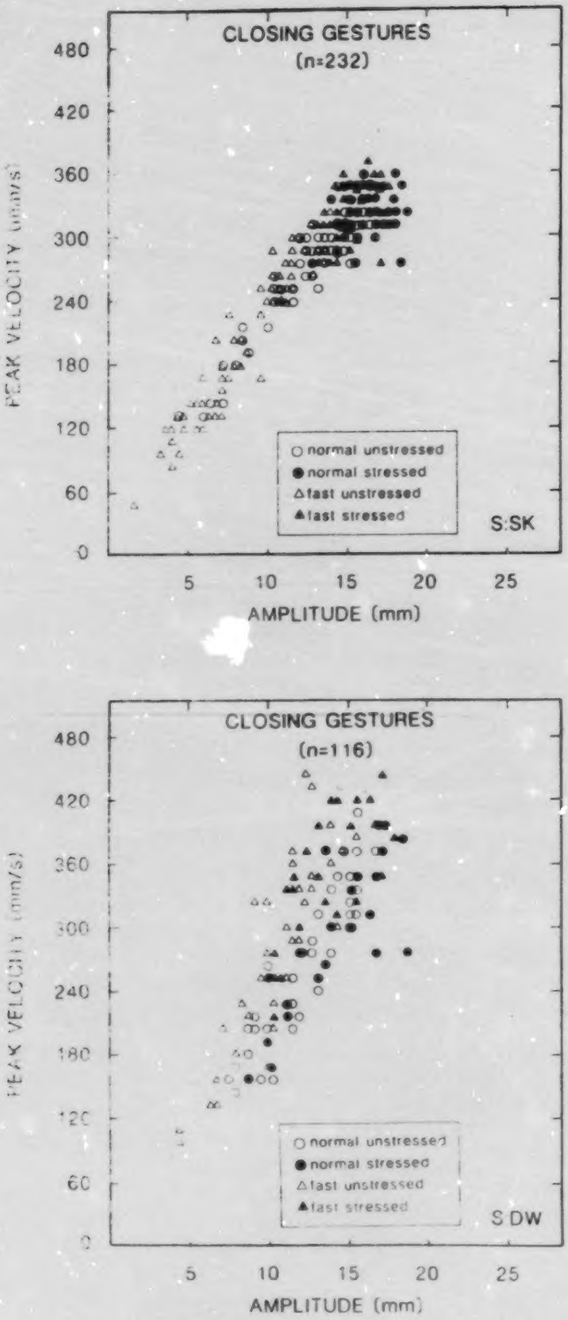


Figure 9. Scatter plot of peak velocity versus (valley-to-peak) amplitude (lower lip) of each subject's closing gestures associated with the vowel-consonant portion of the syllable. Legend specifies conditions.

Table 5

Linear correlations (r) and regression slopes (m) of peak velocity-displacement relationship across rate and stress transformations (/ba/)

A. Opening Gestures

		Stressed		Unstressed	
		m	r	m	r
Normal	DW	19.02	.94	18.02	.90
	SK	9.29	.53	13.19	.79
Fast	DW	14.18	.82	17.02	.87
	SK	10.78	.48	15.68	.94

B. Closing Gestures

		Stressed		Unstressed	
		m	r	m	r
Normal	DW	20.44	.82	20.19	.79
	SK	3.04	.25*	18.00	.95
Fast	DW	20.25	.74	27.39	.87
	SK	13.20	.68	20.92	.95

All r's except those marked by an asterisk are significant at $p < .01$ or greater

It is now generally recognized that many features of single dimensional movements in discrete targeting tasks can be generated by second-order, linear models whose parameters include damping, stiffness, and rest angle (cf. Bizzi, 1980; Cooke, 1980; Fel'dman & Latash, 1982; Kelso & Holt, 1980 for reviews). In short, the limb exhibits behavior qualitatively similar to a damped mass-spring system for these tasks (Fel'dman, 1966). Such systems are intrinsically self-equilibrating in the sense that the "endpoints" or "movement targets" are achieved regardless of initial conditions. In normal and deafferented animals, for example, it has been shown that desired head (Bizzi, Polit, & Morasso, 1976) and limb positions (Polit & Bizzi, 1978) are attainable without starting position information even when the limb is perturbed on its path to the goal. Similarly, Kelso (1977) demonstrated that finger localization ability is not seriously impaired in functionally deafferented humans, or individuals with the metacarpophalangeal joint capsule surgically removed, in spite of changes in initial conditions or unexpected perturbations (Kelso & Holt, 1980; see also Kelso & Tuller, 1983, and Tye, Zimmermann, & Kelso, 1983, for evidence in speech). Closed-loop notions that rely on peripheral feedback break down in the face of such evidence. Further, kinematic variables need not be controlled explicitly. In a dynamic system like a damped mass-spring (or point attractor, Abraham & Shaw, 1982), kinematic variations in displacement, velocity, and acceleration occur as a result of the specified parameter values, and sensory "feedback" in its conventional form is not required. *Nor, importantly, is duration a controlled variable* (see Section IIB).

For sustained, stable cyclic movements of dissipative systems the appropriate dynamic regime is a limit cycle (or *periodic attractor*, Abraham & Shaw, 1982). In such systems, the same orbit is achieved regardless of initial conditions or temporary perturbations. In the absence of imposed perturbations, such systems can display near-sinusoidal steady-state motions that may be treated as if they were generated by simpler nondissipative mass-spring dynamics. As mentioned earlier, a constant slope in the relationship between each gesture's peak velocity and displacement for a given set of gestures indicates that the gestures are perfectly isochronous. With regard to an hypothesized underlying linear (harmonic) or nonlinear (anharmonic) mass-spring model, the Vp-d slope is indicative of the stiffness over the range of gestural displacements examined. Roughly speaking, a constant Vp-d slope for a given gestural subset implies that the average mass-normalized stiffness (K^*_{av}) of the spring functions underlying the gestures is the same across the observed range.* Recently, Ostry, Keller, and Parush (1983) have shown in a study of tongue dorsum movement that the slope of the Vp-d relation varies systematically with stress, but less so as a function of rate. In their data, particularly for opening gestures, the slope of the relationship was greater for unstressed than stressed gestures, suggesting to them that the tongue muscle system was actually *stiffer* in the unstressed environment (see also Laferriere, 1982, for evidence leading to the same conclusion). More recently, observations of tongue dorsum kinematics as a function of rate, vowel (/u/, /o/, and /a/), and consonant (/k/ and /g/) have been interpreted as indicative of an underlying mass-spring control regime with *constant* linear stiffness for a given gesture (Ostry & Munhall, 1984; see also Munhall & Ostry, 1984).

Our data also suggest that unstressed gestures are characterized by greater stiffness (K^*_{av}) values (as revealed in Table 5 by the slopes of the Vp-d relations and the phase portraits) than stressed ones. This is apparent in three out of four cases for both opening and closing gestures (Table 5). Interestingly, we show also that the Vp-d slopes for closing gestures (again with a single exception) are greater than those for opening gestures, particularly for unstressed syllables. Like the Ostry et al. (1983) tongue data, the rate effects on the slope of the Vp-d relationship are less clear cut. In only five of eight possible cases, slope increases as a function of rate. With one notable exception, however, in which a fourfold increase in slope occurs, slope changes between fast and normally produced gestures are fairly small.

Although the data in general suggest that stiffness (K^*_{av}) is a key system parameter, a comparison of the Vp-d slope data (which indexes K^*_{av}) and the displacement data shown in Tables 2 and 3 reveals that stiffness is not constant for movements of different displacements within a given stress condition (see also Ostry & Munhall, 1984). In fact, stiffness changes invariably with displacement both within and across stress categories.

3. The acceleration-displacement function.

There are at least two possibilities that can account for the observed change in stiffness (K^*_{av}) as a function of displacement. One is that a linear spring function holds, for which spring force equals $-kx$ and for which different values of linear stiffness, k, are elected for, say, shorter amplitude, unstressed gestures than larger amplitude, stressed gestures. An alternative notion is that during reiterant speech the jaw-lip system behaves like a soft, nonlinear mass-spring system where, for example, spring force equals $-kx + ex^3$, with k and e denoting linear and cubic stiffnesses, respectively (cf. Jordan & Smith, 1977; Kelso, Putnam, & Goodman, 1983). For such springs, the net stiffness decreases nonlinearly with deviation from the equilibrium position. Hence, shorter amplitude gestures, involving relatively small deviations from equilibrium, are characterized by higher average stiffnesses over the course of the movement than larger amplitude gestures (see also Footnote 4). This second hypothesis is presaged on the assumption that all the motions we have observed arise from a single underlying nonlinear spring function with constant linear and cubic stiffness coefficients. Since a gesture's linear stiffness coefficient is indexed by the slope of the acceleration-displacement function near the gesture's midpoint (corresponding roughly to its equilibrium position), we can distinguish between these foregoing alternatives.

The acceleration data of the lower lip-jaw combination were obtained by velocity differentiation and smoothed over a 25-ms interval (see Section I). Linear instantaneous, mass-normalized stiffness, K^*, was estimated using a computer routine that first found the midpoint of a given opening or closing gesture and then obtained the position (x) and acceleration (X) coordinates of the data sample to each side of the midpoint. This procedure allowed us to compute the slope around the hypothesized equilibrium position. If K^* is unchanged across conditions the slopes should be statistically equal. Thus if the data lie on a single spring function (linear or nonlinear) K^* should be identical close to the movement's midpoint. Different slopes of the x, X function, however, would suggest separate spring functions with distinct linear stiffness components.[5]

Figure 10 (inset) shows how K^* was estimated and also an example of the acceleration-displacement differences between the opening and closing gestures of a stressed versus an unstressed syllable, the fourth and fifth syllables (underlined) of the reiterant versions of "There is ac-cor-ding to legend" (normal rate, SK). Differences in slope are apparent, with the shorter amplitude, unstressed gestures displaying greater K^* values than the longer amplitude, stressed gestures.

Statistical analysis bears this picture out. The mean estimated K^* and its standard deviation are provided for each subject and each gesture type as a function of conditions in Table 6. Stressed gestures as a class have lower K^* values than unstressed gestures, $Fs = 9.38$ and 192.13, $ps < 0.01$ for DW and SK, respectively. Subject DW displays a gesture type main effect, $F = 19.16$, $p < 0.0001$ with K^* significantly greater for closing than opening gestures. Additionally, for SK there is a gesture type X stress interaction, $F = 20.39$, $p < 0.0001$, and also a gesture X stress X rate interaction, $F = 4.70$, $p < 0.04$. A simple main effects analysis of these interactions revealed that for SK: a) K^* is greater for unstressed gestures than stressed gestures for both opening, $F = 168.85$, $p < 0.0001$, and closing gestures, $F = 43.67$, $p < 0.0001$;

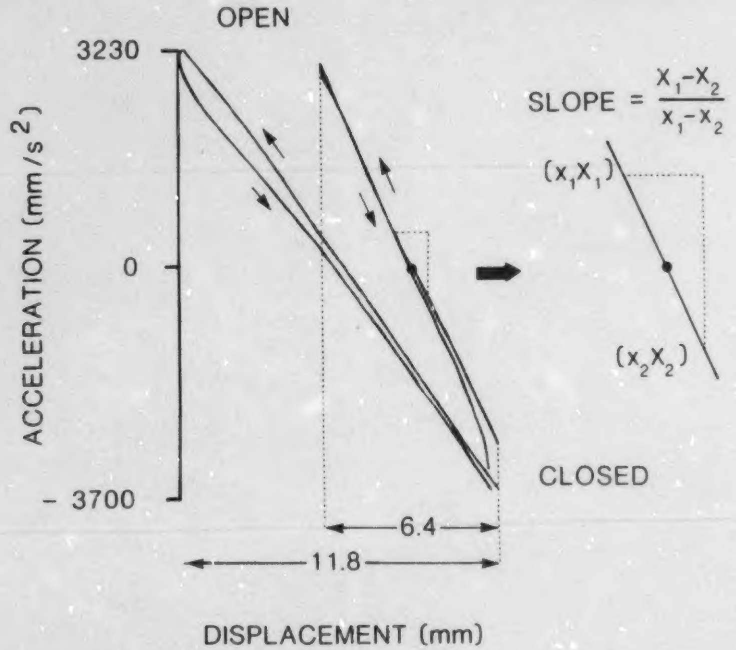


Figure 10. Acceleration versus displacement from rest position for the opening and closing gestures associated with a stressed and unstressed syllable (see text). The smaller displacements and steeper slopes correspond to the unstressed gestures. The opening gestures start at the bottom right; closing gestures start at top left.

Table 6

Estimated stiffness (K*) in units of acceleration per unit displacement across rate and stress transformations. Standard deviation is in parentheses.

A. Opening Gestures

		Stressed	Unstressed
Normal	DW	1443 (405)	1703 (561)
	SK	1781 (342)	2413 (703)
Fast	DW	1931 (836)	2336 (787)
	SK	1803 (427)	3555 (1211)

B. Closing Gestures

		Stressed	Unstressed
Normal	DW	1854 (364)	1889 (496)
	SK	1981 (338)	2308 (344)
Fast	DW	2409 (378)	2633 (452)
	SK	2193 (337)	3078 (880)

b) K* is greater for stressed gestures in the closing phase than the opening phase, $\underline{F} = 8.80$, $\underline{p} < 0.004$; and c) K* is greater for unstressed gestures in the opening phase, $\underline{F} = 12.17$, $\underline{p} < 0.0006$, particularly at the fast speaking rate (see Table 6).

The effect of rate is highly significant for both subjects. In all the cells of similar conditions, K* is greater at the faster speaking rate than it is in syllables produced at a conversational pace, $\underline{F} = 69.43$, $\underline{p} < 0.0001$ (DW) and $\underline{F} = 90.80$, $\underline{p} < 0.0001$ (SK). Subject SK also reveals a stress X rate interaction, $\underline{F} = 41.78$, $\underline{p} < 0.0001$, although DW does not, $\underline{F} = 1.22$, $\underline{p} > 0.05$. For subject SK: a) at both rates, K* is greater in unstressed than stressed gestures, $\underline{F} = 27.39$, $\underline{p} < 0.0001$ (normal) and $\underline{F} = 206.55$, $\underline{p} < 0.0001$ (fast); and b) only in unstressed gestures and in stressed closing gestures, however, is K* greater for fast than for normal speaking rates (see Table 6).

These data correspond rather well to the peak velocity-displacement findings discussed in the Section IIC3. The present acceleration-displacement results, however, afford an additional conclusion, namely, that linear mass-normalized stiffness (K*, estimated around the equilibrium point of the motion) is not the same for short amplitude, unstressed gestures as it is for large amplitude, stressed gestures. In short, different stress categories are characterized by different K* values. A similar conclusion applies to rate changes. In all the cells from comparable conditions shown in Table 6, faster speaking rates are accompanied by higher estimated K* values, and, as we reported in Section IIC1, smaller displacements. Thus although a constant linear stiffness model is a reasonable first approximation, it does not handle all of the kinematic variations in our data that are induced by stress and rate. For whatever reasons, no doubt in part linguistic, linear stiffness is modulated according to the stress (or amplitude) of the gesture. Increasing stiffness for unstressed (shorter amplitude) gestures may be a way for the English language, conventionally classified as stress timed, to differentiate its stress categories. Interestingly, recent theorizing in speech perception argues for a perceptual metric that is closely tied to articulatory dynamics (e.g., Summerfield, 1979; Studdert-Kennedy, 1983). The notion, based in part on studies of visual perception (e.g., Runeson & Frykholm, 1981) is that perception of events is not simply based on surface kinematics, but on the underlying relations among dynamic parameters that govern such events. The present findings, in showing a clear relation between stress and linear stiffness values, provides an initial grounding for these speculations. The data also show that faster speaking rates are associated with higher estimated linear stiffnesses, though, like the Ostry et al. (1983) tongue data, the rate effects are not quite as consistent.

D. Summary and preliminary dynamic modeling

To summarize, the present data offer insights into both the similarities and differences in our subjects' articulatory behavior. The movements of both subjects can be assumed to emerge from the same underlying dynamic organization. That is, a periodic attractor (limit cycle) control regime can capture the forms of motion produced by both subjects. The slopes of the peak velocity-displacement and the acceleration-displacement functions point to linear mass-normalized stiffness, K*, as a key dynamic parameter. The subjects differ, however, in the degree to which estimated K* and overall gestural displacement are coupled across movement conditions. Subject SK shows an inverse relation between stiffness and displacement for opening ($r = -0.77$) and closing ($r = -0.73$) gestures. Thus larger (smaller) amplitude motions that accompany stressed (unstressed) gestures and normal (fast) rates are associated with lower (higher) stiffness. For DW, however, the correlation between estimated stiffness and displacement (like her displacement-movement time relation) is low (-0.18 for opening and -0.25 for closing gestures), perhaps because of the reasons discussed in Section IIC1. In short, the "strength" of the constraint between K* and displacement may be a useful way to conceptualize between-subject differences.

The present findings can be couched conveniently within a recent dynamic modeling and computer simulation framework developed for multiarticulator systems by Saltzman and Kelso (1983). Briefly, the unique feature of this approach is that invariant dynamical equations of motion are established functionally (i.e., at an abstract task level of movement description) for the particular end effectors directly involved in the task's accomplishment. For example, Saltzman and Kelso (1983) demonstrate that a constant set of dynamic parameters defined for a given task, e.g., a hand reaching for a target, can be used to specify context- (task and posture) dependent patterns of change in the articulator-level dynamic parameters (e.g., joint stiffness, damping, and equilibrium points of shoulder, elbow, and wrist). Among other advantages the approach allows for task-specific trajectory shaping (e.g., Bizzi, Acornero, Chapple, & Hogan, 1982) and the compensatory behaviors typically involved in speech (e.g., Folkins & Abbs, 1975; Kelso, Tuller, V.-Bateson, & Fowler, 1984).

At a recent conference, Browman, Goldstein, Kelso, Rubin, and Saltzman (1984) reported that the task-dynamic approach can be fruitfully applied to understanding speech organization. For example, we have used the average values of amplitude and duration from the present data (for stressed and unstressed gestures at a particular rate) to estimate the dynamic parameters (equilibrium positions and stiffnesses) in a functional mass-spring model for the control of lip aperture defined by the vertical distance between the upper and lower lip. These lip aperture parameters remain invariant throughout a given lip opening or closing gesture, and during each gesture are transformed into contextually varying patterns of dynamic parameters at the articulatory level (upper lip, lower lip, and jaw degrees of freedom as defined in the Haskins Laboratories' software articulatory synthesizer; Rubin, Baer, & Mermelstein, 1981). Thus inserting our empirically estimated dynamic parameters for lip aperture into the task-dynamic model, we can generate sets of simulated articulator trajectories associated with lip opening and closing. Figure 11 illustrates simulated time series and phase plane trajectories for the resultant vertical motion of the lower lip during a reiterant bilabial task with simple alternating stress.

In these simulations, the equilibrium position for a given cycle (closure-to-closure) is specified at the onset of the opening gesture and is located halfway between the maximum opening position and the (relatively fixed) closure position. However, because closing gestures are faster than opening gestures (compare Tables 2 and 3) stiffness is specified twice during the cycle: once at the start of the opening gesture and once at the start of the closing gesture. Although the present example simply shows an alternating stress pattern, clearly this procedure can be executed on a syllable-by-syllable basis. Although the model is presently undergoing refinement (e.g., to incorporate fully limit cycle dynamics), Browman et al. (1984) have used the displacement-time data shown in Figure 11 as input parameters to an articula-

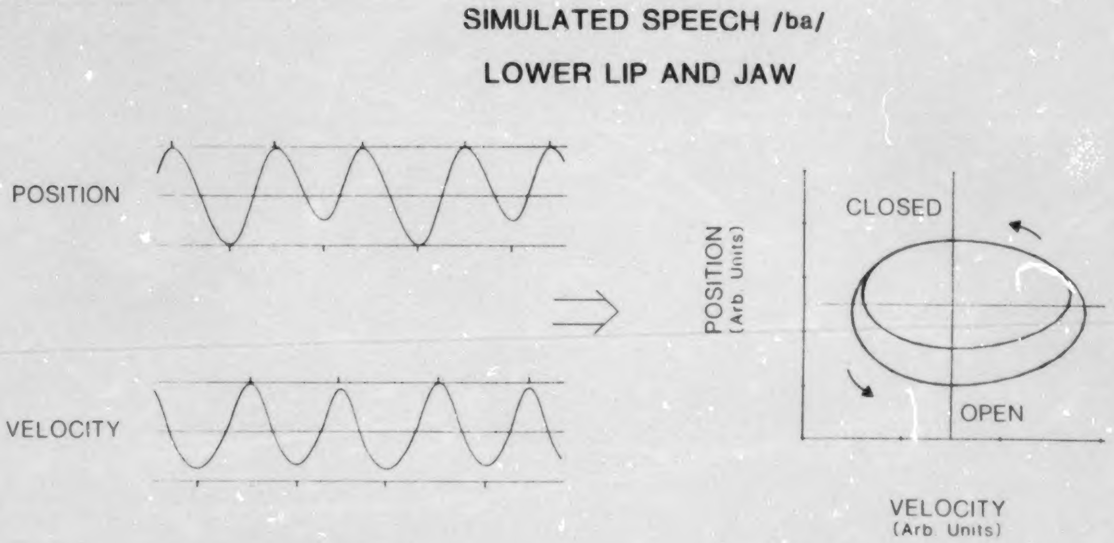


Figure 11. Computer simulation of resultant lower lip position and velocity time series (left) and corresponding phase portraits for a pattern of reiterant, alternating stress syllables.

tory synthesizer with promising acoustic and perceptual results. The point here, however, is that the simulation illustrates how articulatory trajectories can be generated from a simple specification of dynamic parameters without explicit or detailed trajectory planning.

III. Conclusions

The phase portrait methodology introduced in Section IIB, along with a detailed analysis of articulatory kinematics, allow us a window into the hypothesized dynamic structure underlying the production of simple, reiterant syllables. It is popular to propose "time control" as the basis of temporal organization in speech, as if the system somehow had to program and/or keep continuous track of time (e.g., Lindblom, 1963; Lindblom et al., 1984). Different time control schemes, according to this notion, correspond to stress and rate, while other kinematic variables, such as velocity, are computationally derived (cf. Kuehn & Moll, 1976; Laferriere, 1982). In an alternative view, which we have applied here, spatiotemporal pattern arises as a consequence of a dynamic regime in which--at worst--only two articulatory parameters, stiffness and rest position, are specified according to stress and rate requirements. Similar arguments have been proposed for the space-time structure of multidegree of freedom limb movements (Kelso, Putnam, & Goodman, 1983; Kelso, Southard, & Goodman, 1979). The dynamic description captures the forms of articulatory motion observed in our phase portraits across rate and stress conditions. It recognizes in full that articulatory motions *evolve* in time but it undercuts the necessity to *regulate* time as a controlled variable explicitly. Dynamics can provide a grounding for, and a principled analysis of so-called intrinsic timing theories of speech production (Fowler et al., 1980). According to the present findings and supplemented by preliminary modeling, movement time results from an underlying dynamic organization that is specified according to linguistic requirements and that remains invariant throughout the production of a given speech gesture.

References

Abbs, J. H. (1973). The influence of the gamma motor system on jaw movements during speech: A theoretical framework and some preliminary observations. *Journal of Speech and Hearing Research*, *16*, 175-200.

Abraham, R. H. & Shaw, C. D. (1982). *Dynamics--The geometry of behavior*. Santa Cruz, CA: Aerial Press.

Bizzi, E. (1980). Central and peripheral mechanisms in motor control. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland.

Bizzi, E., Accornero, N., Chapple, W., & Hogan, N. (1982). Arm trajectory formation in monkeys. *Experimental Brain Research*, *46*, 139-143.

Bizzi, E., Polit, A., & Morasso, P. (1976). Mechanisms underlying achievement of final head position. *Journal of Neurophysiology*, *39*, 435-444.

Browman, C., Goldstein, L., Kelso, J. A. S., Rubin, P. E., & Saltzman, E. L. (1984). Articulatory synthesis from underlying dynamics. *Journal of the Acoustical Society of America*, *75* (Suppl.1), S22.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic analysis of execution by man of the simplest motor tasks. *Biophysics*, *11*, 766-775.

Fel'dman, A. G., & Latash, M. L. (1982). Afferent and efferent components of joint position sense: Interpretation of kinesthetic illusions. *Biological Cybernetics*, *42*, 205-214.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, *18*, 207-220.

Fowler, C. (1983). Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in sequences of monosyllabic stress feet. *Journal of Experimental Psychology: General*, *112*, 386-412.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press.

Garfinkel, A. (1983). A mathematics for physiology. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, *245*, R455-R466.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), *Speech motor control*. Oxford: Pergamon Press.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. Review of Modern Physics, 47, 67-121.

Haken, H. (1977). Synergetics: An introduction. Heidelberg: Springer-Verlag.

Hogan, N. (1984). An organizing principle for a class of voluntary movements. Journal of Neuroscience.

James, M. L., Smith, G. M., & Wolford, J. C. (1977). Applied numerical methods for digital computation (2nd ed.). New York: Harper & Row.

Jordan, D. W., & Smith, P. (1977). Nonlinear ordinary differential equations. Oxford: Clarendon Press.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. Journal of Neurophysiology, 43, 1183-1196.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.

Kelso, J. A. S., Putnam, C. A., & Goodman, D. (1983). On the space-time structure of human interlimb coordination. Quarterly Journal of Experimental Psychology, 35A, 347-376.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Kelso, J. A. S., & Tuller, B. (1983). 'Compensatory articulation' under conditions of reduced afferent information: A dynamic formulation. Journal of Speech and Hearing Research, 26, 217-224.

Kelso, J. A. S., & Tuller, B. (1984a). Converging evidence in support of common dynamical principles for speech and movement coordination. American Journal of Physiology, 246, R928-R935.

Kelso, J. A. S., & Tuller, B. (1984b). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience. New York: Plenum Press.

Kelso, J. A. S., Tuller, B., V.-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.

Kent, R. D., & Moll, K. (1972). Cinefluorographic analyses of selected lingual consonants. Journal of Speech and Hearing Research, 15, 453-473.

Kiritani, S, Imagawa, H., Takahashi, T., Masaki, S., & Shirai, K. (1982). Temporal characteristics of the jaw movements in the production of connected vowels. Annual Bulletin, Research Institute of Logopedics and Phoniatrics (University of Tokyo), 16, 1-10.

Kozhevnikov, V. A., & Chistovich, L. A. (1966). Rech`, Artikulyatsiya, i vospriyatiye, [Speech: Articulation and perception] (30, p. 543). Washington, D.C.: Joint Publications Research Service. [Orig. published 1965.]

Kuehn, D., & Moll, K. (1976). A cineradiographic study of VC and CV articulatory velocities. Journal of Phonetics, 4, 303-320.

LaFerriere, M. (1982). Stress and tongue blade movement in alveolar VC gestures. Journal of the Acoustical Society of America, 72 (Suppl.1), S104.

Larkey, L. S. (1983). Reiterant speech: An acoustic and perceptual evaluation. Journal of the Acoustical Society of America, 73, 1337-1345.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. Journal of the Acoustical Society of America, 51, 2018-2024.

Lenneberg, E. H. (1967). Biological foundations of language. New York: Wiley.

Liberman, M., & Streeter, L. A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. Journal of the Acoustical Society of America, 63, 231-233.

Lindblom, B. (1963). Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 35, 1773-1781.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag.

Lindblom, B., Lubker, J., Gay, T., Lyberg, B., Branderud, P. & Holmgren, K. (1984). The concept of target and speech timing. Unpublished manuscript. Department of Phonetics, Stockholm University.

Lisker, L. (1975). Phonetic aspects of time and timing. Haskins Laboratories Status Report on Speech Research, SR-47, 113-120.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. Psychological Research, 77, 182-196.

Munhall, K., & Ostry, D. (1984). Ultrasonic measurement of laryngeal kinematics. In I. R. Titze & R. C. Shere (Eds.), Physiology and biophysics of voice. Iowa City, IA: University of Iowa Press.

Nakatani, L. H. (1977). Computer-aided signal handling for speech research. Journal of the Acoustical Society of America, 61, 1056-1062.

Nelson, W. L. (1983). Physical principles of economies of skilled movements. Biological Cybernetics, 46, 135-147.

Nelson, W. L., Perkell, J., & Westbury, J. (1984). Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations. Journal of the Acoustical Society of America, 75, 945-951.

Ohala, J. J. (1975). The temporal regulation of speech. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech. London: Academic Press.

Ohala, J. J., Hiki, S., Hubler, S., & Harshman, R. (1968). Photoelectric methods of transducing lip and jaw movements in speech. UCLA Working Papers in Phonetics, 10, 135-144.

Ohman, S. E. G. (1967). Numerical model of coarticulation. Journal of the Acoustical Society of America, 41, 310-320.

Ostry, D. J., & Munhall, K. (1984). Control of rate and duration in speech. Journal of the Acoustical Society of America, 76.

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 622-636.

Parush, A., Ostry, D. J., & Munhall, K. G. (1983). A kinematic study of lingual coarticulation in VCV sequences. Journal of the Acoustical Society of America, 74, 1115-1125.

Perkell, J. S. (1969). Physiology of speech production: Results and implications of a quantitative cineradiographic study. Cambridge, MA: MIT Press.

Pike, K. (1945). Intonation of American English. Ann Arbor, MI: University of Michigan Press.

Poincaré, H. (1899). Les méthodes nouvelles de la méchanique celeste, Vol. III.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. Science, 201, 1235-1237.

Rosenbaum, D. A., & Patashnik, O. (1980). Time to time in the human motor system. In R. S. Nickerson (Ed.), Attention and performance VIII. Hillsdale, NJ: Erlbaum.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. Journal of the Acoustical Society of America, 70, 321-328.

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. Journal of Experimental Psychology: Human Perception and Performance, 7, 733-740.

Saltzman, E. L., & Kelso, J. A. S. (1983). Skilled actions: A task dynamic approach. Haskins Laboratories Status Report on Speech Research, SR-76, 3-50.

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, J. S., & Quinn, J. T. Jr. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. Psychological Review, 86, 415-451.

Stone, M. (1981). Evidence for a rhythm pattern in speech production: Observations of jaw movement. Journal of Phonetics, 9, 109-120.

Studdert-Kennedy, M. (1983). On learning to speak. Human Neurobiology, 2, 191-195.

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. Journal of Speech and Hearing Research, 16, 397-420.

Tuller, B., Harris, K., & Kelso, J. A. S. (1982a). Stress and rate: Differential transformations of articulation. Journal of the Acoustical Society of America, 71, 1534-1543.

Tuller, B., Harris, K., & Kelso, J. A. S. (1982b). Articulatory control as a function of stress and rate. Haskins Laboratories Status Report on Speech Research, SR-71/72, 81-88.

Tuller, B., & Kelso, J. A. S. (1984). On the relative timing of articulatory gestures: Evidence for relational invariants. Journal of the Acoustical Society of America, 76, 1030-1036.

Tuller, B., Kelso, J. A. S., & Harris, K. (1982). Interarticular phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 8, 460-472.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1983). Further evidence for the role of relative timing in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 829-833.

Tye, N., Zimmerman, G., & Kelso, J. A. S. (1983). Compensatory articulation in hearing-impaired speakers: A cinefluorographic study. Journal of Phonetics, 11, 101-115.

Footnotes

[1] For many examples of complex, multicomponent systems in physics, chemistry, and biology, whose cooperative behavior can be described by a small set of dynamic or "order" parameters, see Haken (1975, 1977). For examples in speech and other biological motions, see Kelso and Tuller (1984b).

[2] The field of qualitative dynamics has a rich history dating back to Poincaré (1899) (see Abraham & Shaw, 1982; Garfinkel, 1983). In this vein, we combine geometry and dynamics to reflect our concern with the forms of

articulator motion (indicated by patterns of displacement, velocity, and acceleration over time) that are created by a functionally defined dynamical organization (e.g., point attractor or periodic attractor dynamics).

[3]Note that the plotting convention here is not the one typically used in dynamics, which plots position on the horizontal axis and velocity on the vertical axis. Since the displacements measured here are vertical, not horizontal, we have simply switched the axes to conform to the behavior of the lip-jaw system and to facilitate visualization of the process.

[4]Both linear and nonlinear mass-spring systems can display near sinusoidal cyclic motions whose observed peak-to-peak period $T = 2\pi/\Omega_c$, where Ω_c denotes the observed angular frequency for the cycle. For systems with constant parameters, the peak-to-valley duration ($D_p = \pi/\Omega_p$) and the valley-to-peak duration ($D_v = \pi/\Omega_v$) are equal and, consequently, $T = D_p + D_v = 2D_p = 2D_v$, and $\Omega_c = \Omega_v = \Omega_p$. More generally, in cases where motion during each half-cycle is near-sinusoidal but of different duration we have $\Omega_c = 2(\Omega_v\Omega_p/[\Omega_v+\Omega_p])$. A linear undamped mass-spring system (harmonic oscillator) may be characterized by the following equation of motion with constant parameters:

$$mX + k\Delta x = 0,$$

where m=mass, k=linear stiffness, $\Delta x = (x - x_0)$ with x_0 = rest position; and x and X represent position and acceleration, respectively. Such systems display cyclic motions with period $T = 2\pi/\Omega_c$ where $\Omega_c = \omega_0 = (k/m)^{\frac{1}{2}}$, and ω_0^2 (denoted K^* in Section IIC3) defines the mass-normalized linear stiffness of the system. Due to system linearity, the instantaneous system stiffness is independent of displacement and, hence, both the instantaneous and the "average" stiffness of the system for motion cycles of different amplitudes are simply equal to k. Normalizing with respect to mass, we see that the average mass-normalized stiffness described in the text, K^*_{av}, is simply k/m ($= \omega_0^2 = \Omega_c^2$) for linear mass-spring systems. Additionally, the peak velocity (V_p) for harmonic oscillators is $\omega_0 A$, where A denotes the maximum displacement from the rest position during cyclic motion. Consequently a plot of Vp versus A for different amplitude cycles of a given linear oscillator shows a straight line whose slope equals ω_0. Thus, for a given linear mass-spring system the Vp-A slope is equal to $\omega_0 = \Omega_c = (K^*_{av})^{\frac{1}{2}}$ and is constant across the entire displacement range. For undamped mass-spring systems with nonlinear stiffness functions (anharmonic oscillators), however, the average mass-normalized stiffness, K^*_{av}, for motion cycles of different amplitude is not so simply related to the system's instantaneous stiffness. For example, for a soft nonlinear spring (cf. Jordan & Smith, 1977) the equation of motion is:

$$mX + k\Delta x - e\Delta x^3 = 0,$$

Where e = cubic stiffness, and all other terms are as in the linear case. Here, the system's instantaneous stiffness does not equal k but is a nonlinearly decreasing function of the amplitude of motion. Thus, the system's K^*_{av} will vary for cycles of different amplitude with K^*_{av} decreasing for increasing amplitudes. Additionally, the plot of Vp versus A for different cycles will have a slope that is a decreasing function of amplitude, unlike the linear systems described above. Yet, like these linear systems, the Vp-A slope is still proportional to $(K^*_{av})^{\frac{1}{2}}$.

[5]We had two concerns about the derivation of acceleration. First, we wanted to ascertain how the elected smoothing window changed the values of the central portion of the trajectories where the slope of the acceleration-displacement function was calculated in this study. Second and relatedly, we

wished to ascertain how our derived (smoothed and differentiated) acceleration compared with actual accelerometric data. Two independent methods were used to examine the effects of numerical smoothing and differentiation on the acceleration data; in both cases, the effects were small. (1) To assess the effects of smoothing on the reported data itself, the x,X slope was calculated for a subset of subject SK's reiterant productions (16 opening gestures representative of the overall stress/rate distribution) at four degrees of smoothing: 15 ms, 25 ms, twice at 25 ms (the condition used in the text), and once at 25 ms and again at 45 ms. Increased smoothing reduced mean slope values, $F(3,60) = 3.48$, $p < .05$, but did not change the pattern relative to overall gesture displacement, $F(6,56) = .37$, $p > .1$. (2) The influence of the double differentiation (i.e., acceleration derived from position) and concomitant smoothing procedures was tested at a movement frequency (5 Hz) comparable to that used by speakers in the present study (see Table I), by comparing the output of an Entran accelerometer (model EGC-240-10D) to the second derivative of position output smoothed twice at 25 ms. Taking into account the gain reduction induced by smoothing (see above), we found the average, absolute difference between transducer (unsmoothed) and numerical (smoothed twice at 25 ms) acceleration to be less than 5 percent of the range of measured movements. The cross-correlation between the raw, unsmoothed and the smoothed, derived signal was $r = .98$. *Note*: Not all the x,X functions approximated straight lines as closely as those illustrated in Figure 10. Some were S-shaped ("hooked" at displacement extrema). However, our smoothing procedures did not remove the "hooks." More important, our estimates of K^* were not affected by the presence of such "hooks."

A THEORETICAL NOTE ON SPEECH TIMING*

J. A. S. Kelso,† Betty Tuller,†† and Katherine S. Harris†††

We wish to make a few brief comments on the commentators' remarks and then introduce a representation of interarticulator timing in which time itself is not explicitly involved. To show that such a representation is valid will require a recasting of what data there are on relative timing (e.g., those discussed by Löfqvist) into a geometrical, phase portrait description of articulator trajectories. We have begun to do this. The phase portrait captures the forms of motion caused by an underlying dynamic organization (Abraham & Shaw, 1982; Kelso & Bateson, 1983; Kelso, Holt, Rubin, & Kugler, 1981; Saltzman & Kelso, 1984), in which time as we traditionally measure it (e.g., as duration, latency) is nowhere to be seen. We believe that certain advantages for understanding speech motor control and developing articulatory models accrue immediately from this perspective. But first some comments.

1. Our paper presents a systematic set of data in favor of relative timing among pertinent articulatory gestures. It is an effort to understand the behavior of an articulatory system that is stable across linguistically meaningful transformations. Relative timing, as we propose it, is simply an index of a temporally stable state. It should not be considered as mandatory (Perkell) or necessarily inherent (Clark & Palethorpe). Clark and Palethorpe (this volume) set up a binary distinction (acoustic versus articulatory) that is not one we have ever subscribed to.

2. Our paper identifies the timing of articulatory gestures associated with consonants relative to those associated with flanking vowels. Because strictly speaking, the data presented by Perkell (though interesting) do not pertain to this issue, brevity precludes any extended commentary. The reader should be aware, however, (a) that other, very different accounts exist of coarticulatory effects of the kind discussed by Perkell (e.g., Bell-Berti & Harris, 1979); (b) Clearly many variables are involved in any account of speech production as Perkell notes. To say, however, that the variability in observable output (e.g., trajectories) is accounted for by the variability in

---

*Authors' response to comments by P. F. MacNeilage, A. Löfqvist, J. E. Clark, S. Palethorpe, and J. S. Perkell, on a paper by K. S. Harris, B. Tuller, and J. A. S. Kelso entitled "Temporal invariance in the production of speech." In J. S. Perkell & D. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum, in press.

†Also Departments of Psychology and Biobehavioral Sciences, University of Connecticut.

††Also Cornell University Medical College.

†††Also Graduate Center, City University of New York.

Acknowledgment. This paper and the research discussed herein were supported by NINCDS Grants NS-13617, NS-17778, BRS Grant RR-05596, and ONR Contract No. NC0014-83-C-0083. Thanks to Elliot Saltzman for very helpful suggestions on the manuscript.

programming is circular at best (cf. Kelso, 1981); (c) Abbs' paper (this volume and elsewhere), MacNeilage's writings (e.g., MacNeilage, 1980) and our experiments (Kelso, Tuller, Bateson, & Fowler, 1984; Kelso, Tuller, & Fowler, 1982) all converge in showing that the speech motor control system does not program targets in articulator space (cf. Perkell); and (d) we certainly agree with Perkell that more data are needed, but so are new concepts.

3. In our final remarks we want to return to the theme of our paper, namely, the relative timing among articulatory gestures. We wish to show how, by examining the data using phase plane techniques an entirely different conceptualization for the relative timing finding emerges. We are presently analyzing existing data and conducting new experiments to examine this conceptualization further. Consider the simple case in which the latency (in ms) of onset of upper lip motion for a medial consonant is measured relative to the interval (in ms) between onsets of jaw lowering motions for flanking vowels. These events are displayed in the idealization of Figure 1A, in which the duration of the V1 to V2 cycle is Jd (in ms) and the latency of upper lip motion is Ll (in ms). As we have shown, the two events are highly correlated across rate and stress changes. That is, the lip latency varies systematically with jaw cycle duration plus an intercept value that seems to change across phonetic context and speaker (see Figure 1B). Note that this is a strictly temporal description. One could posit, in this example, that somehow the system is keeping track of the duration of jaw motion such that when a given amount of time has passed, another articulator, say the upper lip, is activated. Such an account of speech or limb movement control is not unusual.

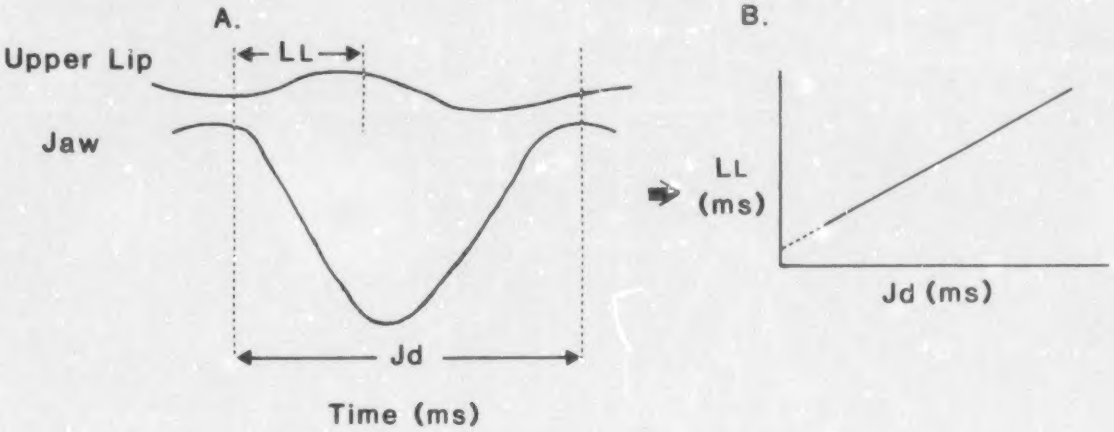
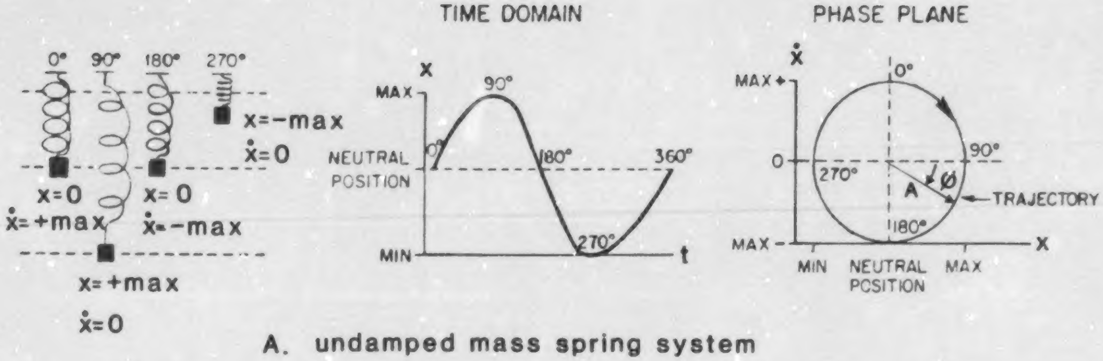


Figure 1. A. An idealized time series description of jaw and upper lip motion. B. The empirical relation between jaw cycle duration, Jd, and upper lip latency, LL (see text for details).

A very different view of these events emerges when the articulatory data are expressed as motion trajectories on the phase plane. Two quantities are needed to do this, the articulator's displacement (x) and its velocity ($\dot{x}$). These quantities may be considered to be the coordinates of a point on the articulator in two dimensional space, the phase plane. As time varies, the point $P(x,\dot{x})$ describing the motion of the articulator moves along a certain path on the phase space. Note that time, although implicit and recoverable from this representation, does not appear explicitly in the phase plane description. For different initial conditions, the corresponding paths will be different, and the set of all possible trajectories constitutes the phase portrait of the system's dynamic behavior. Finally, one can transform the Cartesian $x,\dot{x}$ coordinates into an equivalent polar form described by a phase angle ($\phi=\tan^{-1}[\dot{x}/x]$), and a radial amplitude ($A=[x^2 + \dot{x}^2]^{\frac{1}{2}}$. In discussions below, the phase angle is a key concept in our interpretation of interarticulator timing phenomena.

Figure 2A illustrates the mapping from time domain to phase plane for a motion trajectory generated by a simple, undamped mass-spring system.[1] In a similar fashion, Figure 2B shows the phase plane trajectory for the idealized jaw motion described as a time series event in Figure 1A. In the phase plane, this jaw motion describes a closed orbit, since the jaw goes from closed to open and back to closed in one cycle. Note that, in comparison to Figure 2A, the axes in Figure 2B have been interchanged in order to express pictorially that the jaw moves vertically in space. In the phase plane, one can plot jaw motions during V1V2 intervals of different duration, and can identify the onset of upper lip motion during each cycle with an onset phase angle for that cycle. Our hypothesis is that the phase angle for upper lip onset should be the same across jaw cycles of different shape, i.e., across different rate and stress conditions. Two idealized examples are illustrated in Figure 3. In one, a small orbit is shown, corresponding to a small displacement of the jaw over time. In the other, a larger orbit is shown. The phase angle of upper lip onset, Ø, is predicted to be invariant as shown in the right hand side of Figure 3, though we do not claim it to be the one shown here.[2] Note that the onset of a remote articulator (e.g., the upper lip) is now with reference to the phase angle of another articulator (e.g., the jaw). This angle is therefore a function of the latter articulator's position *and* velocity, not merely its absolute position or velocity. *Moreover, there is no need to posit any kind of time-keeping mechanism or time controller.* In this view, individuals can produce articulatory motions of different durations or magnitudes without affecting the hypothesized regularity in onset phase angle.

To summarize our theoretical points: When representing articulatory motions geometrically on the phase plane, neither absolute nor relative time need be extrinsically monitored or controlled. This fact potentially provides a grounding for, and a principled analysis of, so-called intrinsic timing theories of speech production (e.g., Fowler, Rubin, Remez, & Turvey, 1980; see also Kelso & Tuller, in press). Our view is supported indirectly by demonstrations in the articulatory structures themselves of afferent bases for phase angle information (e.g., position and velocity sensitivities of muscle spindle and joint structures), but not for time-keeping information (e.g., time receptors; cf. Kelso, 1978). It might well be the case that certain critical phase angles provide information for orchestrating the temporal flow of activity among articulators (beyond those considered here) and/or vocal tract configurations. Such phase angles would serve as natural, i.e., dynamically specified, information sources for coordinating speech. Interarticulator



A. undamped mass spring system

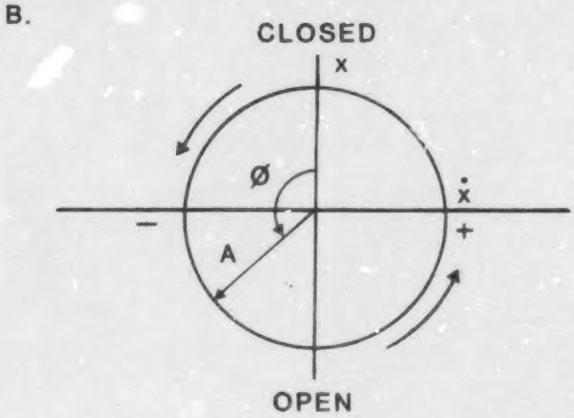


Figure 2. A. The mapping of a simple undamped mass spring motion on to the phase plane. B. The jaw cycle of Figure 1A characterized on a 'functional' phase portrait, i.e., displacement is on the vertical axis and velocity on the horizontal axis. The polar coordinates, the phase angle, ϕ, and the radial amplitude, A, are also shown in the diagrams (see Text for details).

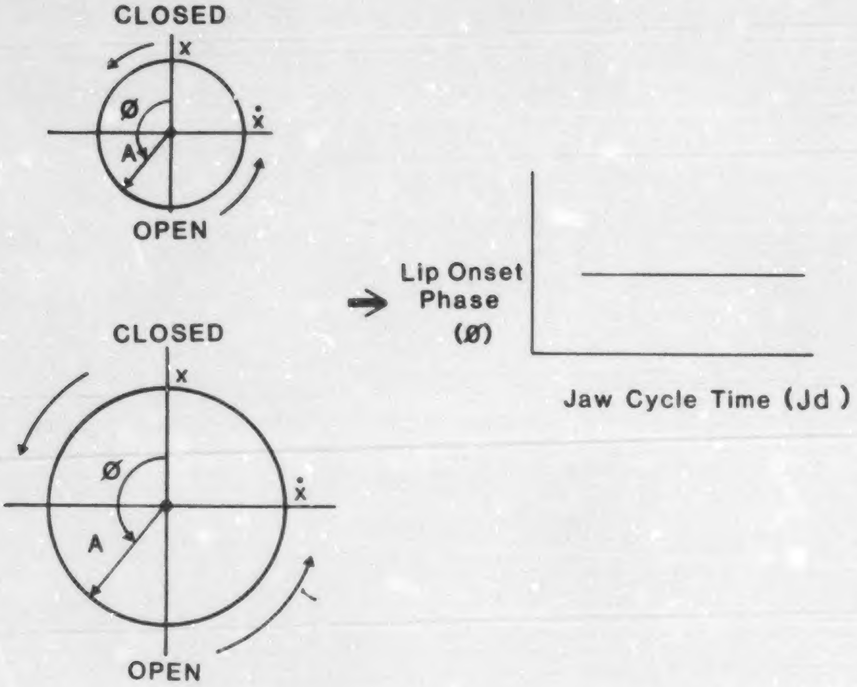


Figure 3. The phase position of the upper lip relative to the jaw cycle for different jaw orbits (see text) and the consequent hypothesized relation (see Footnote 1).

speech coordination thus may be captured better with reference to events that are specified by the system's dynamics than with reference to sets of durational rules. These ideas and others are also currently being explored by dynamic modeling (Saltzman & Kelso, 1984).

References

Abraham, R. H., & Shaw, C. D. (1982). Dynamics--The geometry of behavior. Santa Cruz, CA: Aerial Press.

Bell-Berti, F., & Harris, K. S. (1979). Anticipatory coarticulation: Some implications from a study of lip rounding. Journal of the Acoustical Society of America, 65, 1268-1270.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. New York: Academic Press.

Kelso, J. A. S. (1978). Joint receptors do not provide a satisfactory basis for motor timing and positioning. Psychological Review, 85, 474-481.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Bateson, E.-V. (1983). On the cyclical basis of speech production. Journal of the Acoustical Society of America, 73, S67.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.
Kelso, J. A. S., & Tuller, B. (in press). Intrinsic time and speech production: Theory, methodology, and preliminary observations. In E. Keller (Ed.), Sensory and motor processes in language. Hillsdale, NJ: Erlbaum.
Kelso, J. A. S., Tuller, B., Bateson, E.-V., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.
Kelso, J. A. S., Tuller, B., & Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. Journal of the Acoustical Society of America, 72, S103.
MacNeilage, P. F. (1980). Distinctive properties of speech control. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.
Saltzman, E. L., & Kelso, J. A. S. (1984). Skilled actions: A task dynamic approach. Haskins Laboratories Status Report on Speech Research, SR-76, 3-50.

Footnotes

[1] Note that the jaw motion, though idealized here, does not have to be (and is not usually) sinusoidal. Thus, different relative timing relations among articulators can give rise to the same phase position between articulators and vice versa. The determining feature is the shape of the trajectories (for many more details, see Kelso & Tuller, in press).

[2] To date, we have examined this relationship for two speakers and two phonetic contexts, /babab/ and /bawab/. In each case, the phase angle of the upper lip for the medial consonant relative to the jaw trajectory was unaffected by changes in stress and speaking rate.

ON RECONCILING MONOPHTHONGAL VOWEL PERCEPTS AND CONTINUOUSLY VARYING F PATTERNS*

Leigh Lisker†

Abstract. When a sequence of pictures is presented in rapid succession, the illusion of continuous movement can be created. A continuously varying acoustic signal may, contrariwise, be perceived as a sequence of "still" sounds. Not only is speech perceived as discrete sounds in sequence, but speakers will oblige, especially in the case of stressed vowels, by "citing" them in the form of steady state phonations judged to match auditorily the vowels in their natural contexts. These steady state imitations are adequately characterized by just two numbers, the frequencies of the two lowest vocal-tract resonances. Acoustic analyses of a number of tokens of the English nonsense forms or words [bɛb, dɛd, gɛg, bæb, dæd, gæg] produced in the frame Please pronounce ____ once again by four talkers indicate that there is a within-talker pattern of variation rather different from the variations over speakers reported by Peterson and Barney (1952). Moreover, the variation patterns are different within syllable types, for the same vowel across the contexts examined, and for the two formants. There are differences in the way in which F1 and F2 vary with variation in stop place of articulation and in the voicing of the postvocalic stops. These variations are in some cases of a kind to pose difficulties for the target-plus-undershoot model as the explanation for the variations observed. They are of a magnitude, moreover, that should discourage an attempt to classify vowels automatically on the basis of F1-F2 frequency measurements at a single point on their trajectories and without regard to their context.

The tradition of representing vowel quality acoustically by a point on the plane whose dimensions are the frequencies of the two lowest resonances of the vocal tract is a long one, with its beginning at a time when the analysis of speech and other nonstationary signals was not possible. Instead of normal speech, the objects of analysis were vocalizations produced with a vocal tract held in fixed position over relatively long intervals. Such vocalizations are speech only by courtesy of the fact that they are judged auditorily to match vowels as components of speech events. Considerable attention is still being given these "nonspeech" sounds, whether of vocal tract or machine origin, but for a different reason, namely, that the dynamic nature of speech activity

---

*This paper was presented orally at the 107th meeting of the Acoustical Society of American, Norfolk, VA, 6-10 May 1984.

†Also University of Pennsylvania.

Acknowledgment. This was supported by NICHD grant HD-01994 to Haskins Laboratories.

perturbs what is hypothesized to be the underlying structure of speech events, which we want to think of as sequences of discrete elements, each a complex of features that jointly determine membership in one of a limited set of sound categories. For each category a certain articulatory target and associated acoustic pattern are posited. This target emerges more or less clearly in the so-called null context and in some few others, such as English /h-d/, where disturbing coarticulatory effects are said to be minimal. Contamination by context is both hailed as an essential property of speech and condemned as a confounding (and confounded) impediment to determining a straightforward relation between acoustic signal and linguistic percept.

Different opinions, undoubtedly based on different equally valid observations, have been expressed regarding the scope of contextual perturbation of vowels. Thus Schouten and Pols (1979) found that the Dutch vowels they studied had steady state intervals whose spectral shapes varied little with context, but Lindblom and Studdert-Kennedy (1967) reported that in CVC syllables the formants rarely "reach a steady state," and that under changes in the overall duration of synthetic CVC patterns there are shifts in vowel identification. Presumably these shifts tell us something about variations in natural speech, and specifically about variations in what we call the primary correlates of vowel quality, the F1 and F2 frequencies.

If we accept provisionally the idea that the best place to sample the moving F-pattern to determine F1 and F2 frequencies that will serve as the optimal index of vowel quality is at the point of maximum F1, then it seems to me of some interest to learn how much variation this measure will uncover, and what part of it, if any, may be systematic and attributable to differences in context, and also to see how the magnitude of such context-dependent variation compares with F1 and F2 differences separating distinct vowel categories that are contiguous on the F1-F2 plane. In order to get answers to these questions I recorded the speech of three male speakers of American English varieties that seem very similar phonetically in their [ɛ] and [æ] vowels. The three speakers produced CVC syllables in the carrier sentence Please pronounce ___ once again. Fifteen repetitions of each syllable type were subjected to LPC analysis. A fourth speaker carried out the same recording and analysis procedure as part of his research project for a university seminar course. The data so far analyzed do not allow point-by-point comparison across the four speakers, and a good deal of recorded speech awaits analysis, but already some regularities in the relation between stop context and F1-F2 variations are evident.

In Figure 1 the formant frequencies for 15 tokens of each of the syllables [gɛg, dɛd, bɛb, gæg, dæd, bæb] are represented. Mean values of F1 and F2 are indicated by the location of the intersections of lines whose lengths represent magnitudes of the standard deviations for each formant frequency. It is clear that in the productions of speaker L the first formant is not subject to any major perturbation by context, but that F2 for both the [ɛ] and the [æ] syllables has lower frequencies in labial stop contexts than in dorsal.

Figure 2 shows F1 and F2 plots for the [ɛ] and [æ] syllables in labial and dorsal stop contexts as produced by all four speakers. While there are slight differences among speakers, all show the syllable [gɛg] with lowest F1 and highest F2, and [bæb] with highest F1 and lowest F2. For each vowel and

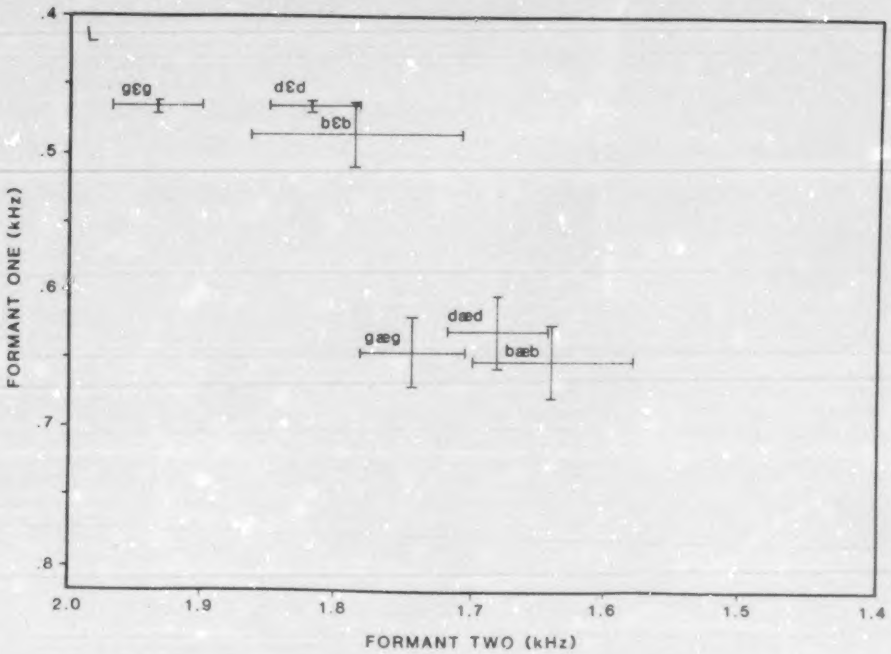


Figure 1. Means and standard deviations of F1 and F2 frequencies, measured at time of maximum F1 frequency, for fifteen tokens of each of six syllable types produced by a single talker L in the frame sentence Please pronounce ____ once again. Lines indicating ± one standard deviation intersect at point representing mean formant frequencies.

each speaker, the dorsal stop environment is reflected in a mean F1 that tends to be lower in frequency and a mean F2 that is clearly higher than in the labial stop context. We may note that for speaker S the syllable [bɛb] is closer to [gæg] than it is to [gɛg], and that [gæg] is closer to [bɛb] than it is to [bæb]. The difference in the apparent effectiveness of F1 and F2 as indices of the [ɛ]-[æ] distinction is made clearer in Figure 3, in which the data from speakers L and W, whose patterns are farthest apart, are plotted together. The two syllable classes can be separated by a boundary at F1 = c.575 Hz. But for F2, while the overall mean value is higher for the [ɛ] syllables, combined speaker and context dependent effects yield some [ɛ] syllable types with rather lower mean F2 frequencies than some [æ] syllables show.

The measurement data analyzed for speaker S include F1 and F2 values of [ɛ] and [æ] vowels in syllables terminated by [p] and [k] as well as [b] and [g] (Figure 4). I expected that the shortening associated with final voiceless stops in place of voiced stops would result in lower maximum F1 frequencies. Instead, as we see here, F1 is higher in [gɛk] than in [gɛg], and likewise higher in the first of the pairs [bɛp]-[bɛb], [gæk]-[gæg], and [bæp]-[bæb]. This effect of final devoicing is somewhat disconcerting, to say the least. If we posit a single target value for all syllables sharing a particular vowel quality, and we further assume that in the case of F1 any failure to reach target is a matter of undershoot and not overshoot, then the

CVC

F1 x F2 at pt. of max. F1

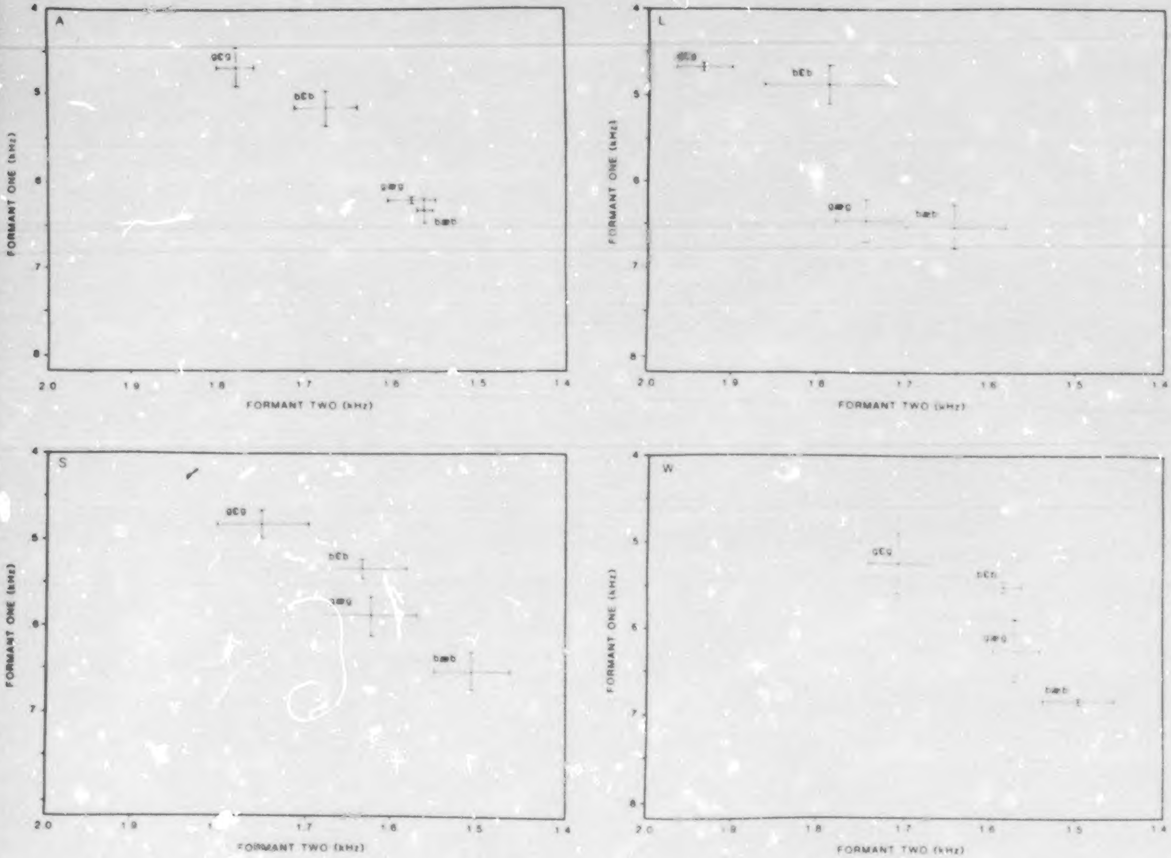


Figure 2. Means and standard deviations of vowels [bɛb, gɛg, bæb, gæg] from four talkers, fifteen tokens of each syllable type per talker.

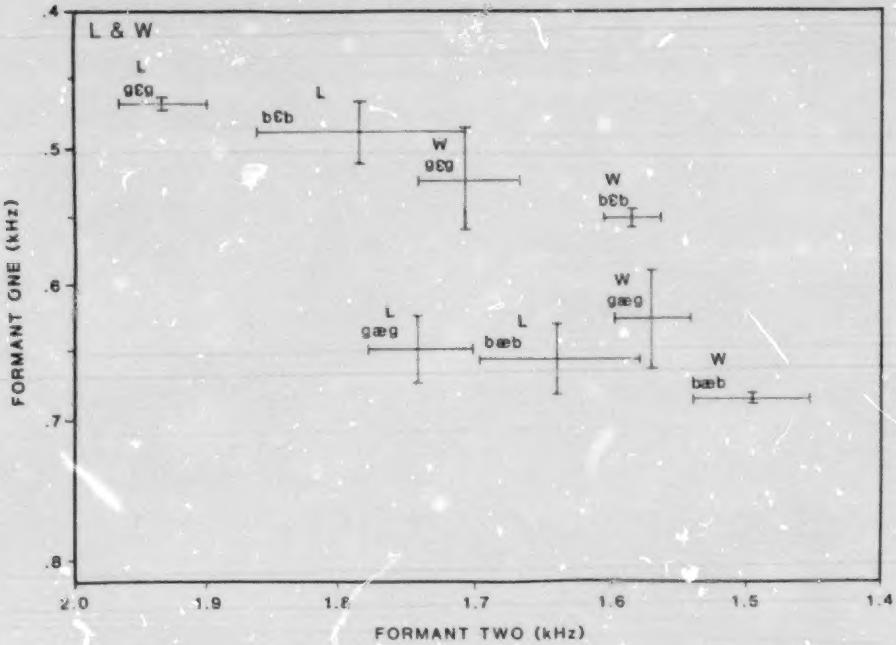


Figure 3. Comparison of two talkers, L and W, showing greatest differences in F1-F2 mean values.

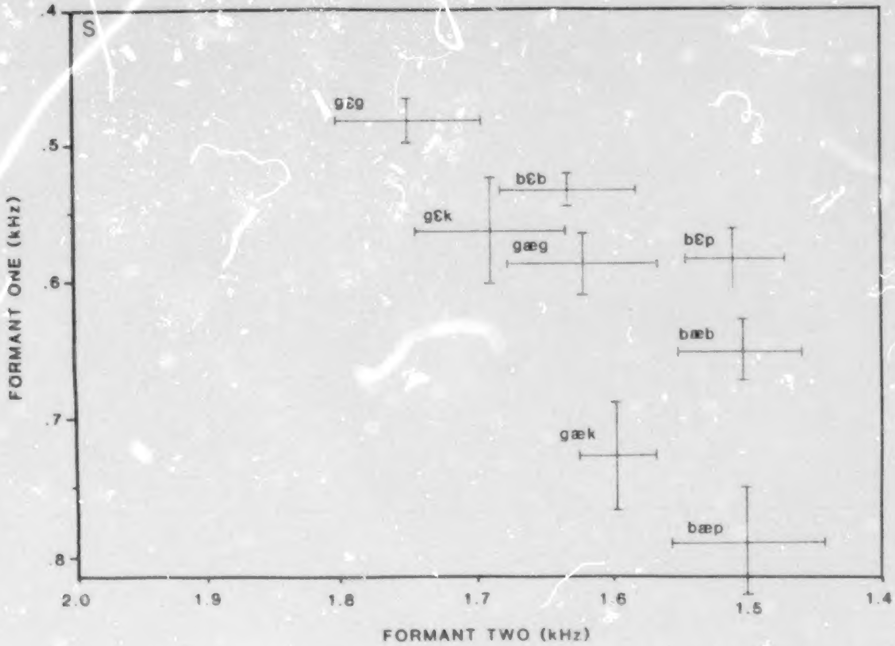


Figure 4. Comparison of F1-F2 frequencies for syllables differing in the voicing of their final consonants, as produced by a single talker, S.

fact that all first formant trajectories in the syllables measured are convex "upward," then a syllable with higher F1 maximum should be closer to target at the point of greatest oral opening. Moreover, a shorter syllable should display greater undershoot, that is, a lower peak F1. Lindblom's studies of vowel reduction (Lindblom, 1963) indicate that the shortening ascribable to a global speedup of articulation or to destressing has this effect. These data fail to conform to the expectation nurtured by the findings of Lindblom (1963) and by Lindblom and Studdert-Kennedy (1967). Can we suppose that the F1 targets for speaker S's [ɛ] and [æ] are more closely approximated in, e.g., [gæk] than in [gæg], that is, that undershoot is incurred with the voicing of the final stop, despite the fact that the duration of the vowel gesture is at the same time slowed? Or can we perhaps entertain the notion that if the offset frequency of F1 is higher before final voiceless stops, this results in a prior raising of F1 that is detectable as early as the point of maximum F1 for the syllable? Perhaps we might entertain the possibility of *overshoot*, particularly if we imagine that [p] and [k] are produced with greater articulatory force than are [b] and [g], unsupported as such an allegation is, and that in consequence the preceding vowels are more energetically articulated, with greater departure from the so-called "neutral" vocal tract shape. The data now on hand need to be augmented before such speculations warrant further discussion.

The data shown in Figure 5 allow us to compare F1-F2 values in symmetrical stop contexts with those found in asymmetrical ones. In the syllables [bɛg], [gɛb], [bæg], and [gæb] the first formants rise and then fall, but the F2 trajectories move in only one direction. We may indeed better suppose that the F2 trajectories *traverse* rather than undershoot any target we might reasonably posit. It appears that in these syllables the F1-F2 values at the point of maximum F1 are more powerfully affected by the postvocalic than by the prevocalic stop. The tendency is for F1 to rise and F2 to descend in the order [gɛg]-[bɛg]-[gɛb] and [gæg]-[bæg]-[gæb]. But this promising regularity is marred by the data for [bɛb] and [bæb], which are not quite nicely placed relative to [gɛb] and [gæb].

The final display (Figure 6) is of data collected to find out how some other vowels with qualities close to those of [ɛ] and [æ] are placed in relation to the latter on the F1-F2 plane. These are the vowels that are represented as [ė] and [ej]. The first has a quality that is distinct from [æ] in the area of the country that includes New York City and Philadelphia: it distinguishes the word *halve* from *have*, for example. The quality of [ej] is usually described as diphthongal: the syllable [bejb] is a pronunciation of the word *babe*. For these two additional syllable nuclei the same effects of labial versus dorsal stop contexts are to be observed. Moreover, it appears that, even though [gejg] and [bejb] are diphthongal as distinct from the others ([ė] is not noticeably diphthongal in L's speech), the single measure of F1 and F2 being tested is as effective in separating [ej] from its closest neighbor [ɛ] as in distinguishing [ɛ] from [ė]. On the other hand, placement in the F1-F2 plane does not well separate [gėg] from [gɛg] and [bɛb]. The effect of substituting [b]s for [g]s as the neighbors of the vowel [ɛ] is greater then replacing [ɛ] by [ė] with the [g-g] context held constant. This fact, if further analysis corroborates it as fact, suggests that the context effects of stop place and voicing can be of a magnitude to put at some risk any procedure of automatic vowel classification that depends on F1-F2 frequency measurements made at a single point in time and without regard to context.

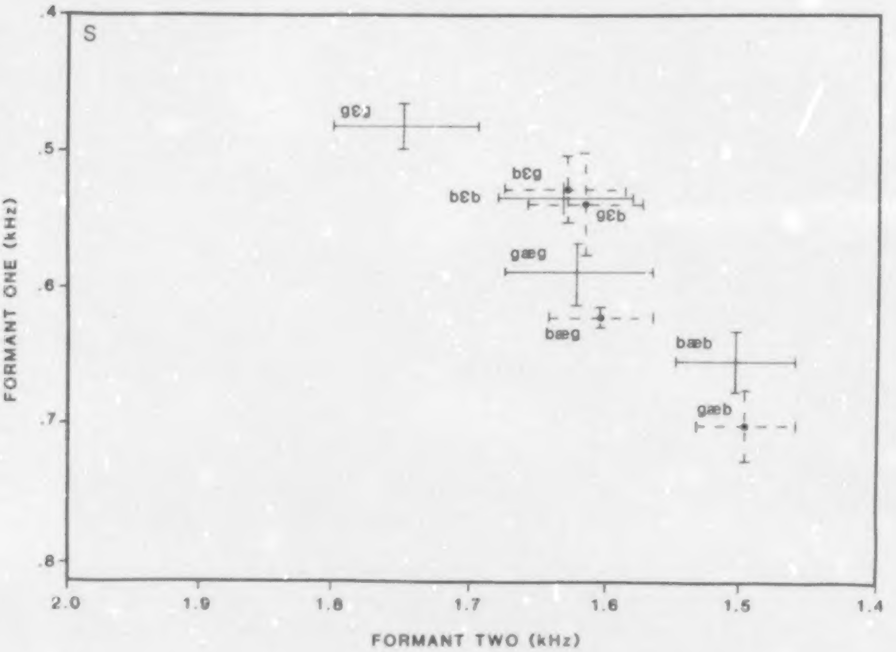


Figure 5. Comparison of F1-F2 frequencies for syllables symmetrical and asymmetrical with respect to their pre- and post-vocalic consonants, as produced by single speaker S.

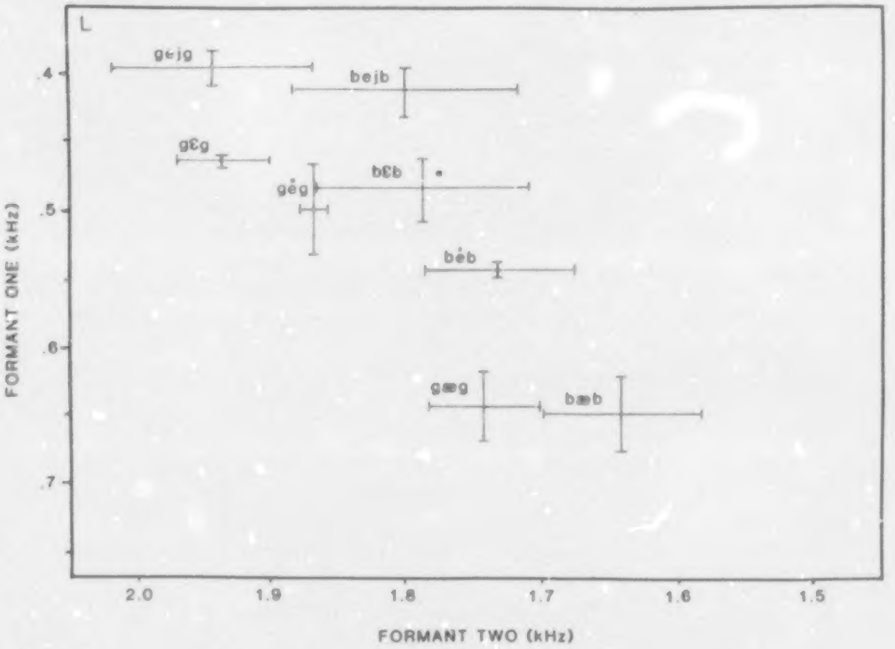


Figure 6. Comparison of F1-F2 values for [ɛ,æ] with the phonologically distinctive vowels [ej] and [ė] in the pattern of speaker L.

To sum up then, F1 and F2 frequencies determined by LPC analysis at the points of maximum first formant frequencies in stop-vowel-stop syllables indicate that for the two "adjacent" vowels [ɛ] and [æ], the maximum F1 frequency is more stable over the set of syllables sharing the same vowel, while F2 frequency varies more with the place of articulation of the flanking stop consonants than it does with the vowel. However, the effect of devoicing the postvocalic stop is more pronounced on F1 than on F2, its magnitude being in fact as great or greater than that interpreted as a shift between [ɛ] and [æ]. These differential effects appear to be similar for syllabic nuclei other than [ɛ] and [æ], in particular the vowel [ė] and the diphthongal [ej].

References

Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 35, 1773-1781.

Lindblom, B. E. F., & Studdert-Kennedy, M. (1967). On the role of formant-transitions in vowel recognition. Journal of the Acoustical Society of America, 42, 830-843.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the identification of vowels. Journal of the Acoustical Society of America, 24, 175-184.

Schouten, M. E. H., & Pols, L. C. W. (1979). Vowel segments in consonantal contexts: A spectral study of coarticulation--Part I. Journal of Phonetics, 7, 1-23.

SYNERGIES: STABILITIES, INSTABILITIES, AND MODES*

E. Saltzman and J. A. S. Kelso†

Nashner and McCollum have addressed the question of whether muscle synergies exist for complex skilled activity, and if so, how they are organized (see also Kelso & Tuller, 1984, and Lee, 1984). The authors argue that muscle synergies exist for postural stability tasks in the form of a small set of discretely represented control entities, and that postural corrective movements of the dynamically continuous musculoskeletal system are organized through the operation of these discrete synergy elements. In this commentary, we make two main points: first, that Nashner and McCollum's arguments are not supported sufficiently by their data (i.e., the data do not allow one to distinguish between their discrete synergy model and other model types). We will describe the sort of data that would be convincing; and second, because Nashner and McCollum stress the "universality and importance of global schemes" for sensorimotor coordination and "principles governing the interactions among elements" that lead to "testable hypotheses" we mention briefly a theoretical framework that is attractive to us (e.g., Kelso, 1984; Kelso & Saltzman, 1982; Kelso & Tuller, 1984; Kugler, Kelso, & Turvey, 1980, 1982) because it treats cooperative behavior in multicomponent systems as an emergent consequence of the systems' underlying dynamics (e.g., Haken, 1975). We feel that this framework can (i) offer a firmer basis for some of Nashner and McCollum's existing experimental observations; and (ii) promote an experimental strategy that would illuminate Nashner and McCollum's hypothesis of region-specific discrete synergies.

Nashner and McCollum describe distinct patterns of EMG bursts in response to distinct patterns of postural perturbation (e.g., vertical or front-back platform translation) in the context of given support conditions (e.g., different platform sizes). Each EMG pattern is characterized by a temporally ordered sequence of bursts within a subset of three agonist-antagonist muscle pairs (ankle, thigh, and trunk muscles). They hypothesize that each such pattern or synergy operates with respect to a corresponding distinct control structure. Each structure controls corrective postural movements within a limited subregion of postural configuration space (e.g., ankle angle vs. hip angle plane), such that when the body is perturbed the associated (fine-tuned) EMG burst pattern will return the body to a balanced posture. In principle,

---

*Slightly revised version of the authors' commentary on target article by Nashner, L. M., and McCollum, G. The organization of human postural movements: A formal basis and experimental synthesis. The Behavioral and Brain Sciences, in press.

†Also Departments of Psychology and Biobehavioral Science, University of Connecticut.

Acknowledgment. Preparation of this paper and some of the research discussed therein was supported by ONR Grant N00014-83-0083 and NIH Grant NS-13617.

however, such synergistic EMG patterns could also be generated by alternative models (e.g., Litvintsev, 1972; Saltzman & Kelso, 1984) in which control laws dependent on task (i.e., maintain balance), support condition, and postural configuration serve to continually specify corrective joint torque vectors that return the body from an unbalanced to a balanced posture. If one defined a further mapping from torque vectors to "muscle element" vectors (e.g., Jerard & Jacobsen, 1980; Saltzman, 1979) for which muscle elements were activated only after inputs exceeded a given threshold, then ongoing corrective torques would be mapped into patterns of discrete EMG bursts in those muscles appropriate for producing the required torques. This sort of control-law model augmented by thresholds for muscle element recruitment should generate consistent "synergistic" patterns of postural EMG in response to given types of destabilizing inputs, without reference to discretely organized synergy control structures. For stabilizing movements initiated from most locations in the postural configuration space, therefore, the above discrete and control-law hypotheses predict qualitatively similar EMG activity patterns. However, the discrete synergy model predicts that there will be certain regions of the configuration space for which the EMG predictions will be different for discrete and control-law models.

For the discrete control hypothesis, partitioning the configuration space into distinct (possibly overlapping) synergy subregions implies that borderlines (or border regions) will be defined between the different control domains (see Nashner & McCollum's Figure 5). Nashner and McCollum's notion implies that the system will behave differently along (or within) these borders than when operating away from the borders. Further, when the postural system adapts from one support condition to another (e.g., from long to short platform lengths) the implication is that the border layout itself shifts correspondingly. Let us focus on the "simpler" adapted case (e.g., repeated trials with short platforms) for which border structure is assumed to be relatively constant. In this instance, the control structures associated with adjacent configurational domains should compete equally at the borders for access to the final common paths of muscular output. There are at least four possible outcomes of such competition: a) opposing effects will cancel each other and no muscle activity will occur; b) competing synergies will be observed simultaneously in a mixture of EMG patterns; c) there will be a repetitive alternation or "jittering" between the EMG patterns of each competing synergy; or d) a totally novel EMG pattern might be observed. Experimental demonstration of any of these patterns near Nashner and McCollum's hypothesized synergy borders in support-condition-adapted subjects would provide strong support for the discrete model, since the control-law model would not behave differently on, near, or away from those borders. These data are lacking, however, or at least have not been presented in the target article. The strongest data offered by Nashner and McCollum in favor of their hypothesis is the sequential mixing of ankle and hip "synergies" during adaptation to suddenly changed platform sizes (see Nashner & McCollum's Figure 7). However, these findings seem equivocal at best given the concomitant shifts in border structure that presumably accompany such adaptation. Therefore, perturbation studies that use adapted subjects and that explore a sufficiently large sample of the postural space could (i) help to identify synergy borders and (ii) constitute a direct test of the discrete synergy model.

The above suggestion exemplifies a general experimental strategy for explicating the cooperative behavior of multicomponent, open, nonlinear systems. A common feature of all such systems is that when control parameters are changed beyond certain critical values, new "modes" or spatiotemporal patterns may appear (for many examples in physics, chemistry, and biology see Haken, 1977, 1983; Prigogine, 1980; Yates, 1982; Yates & Iberall, 1973; for examples in motor behavior, see Cohen, Holmes, & Rand, 1982; Kelso & Tuller, 1984; von Holst, 1973). The beauty of this formulation is that the modes (e.g., synergistic patterns) may themselves be described by a set of dynamical equations derived via transformation procedures from the equations describing the behavior of the original subsystems (e.g., muscle elements). Under the influence of continuous scaling of control parameters, a previously quiescent mode may suddenly become dominant and "capture" the behavior of the subsystems. Such bifurcations result from the competition, as it were, between the "forces" or inputs that are systematically scaled (corresponding, for example, to the direction of platform translation), and the "forces" holding the system together (i.e., the synergistic constraints among muscles).

In Figure 1 we show an example from our own work on cyclic behavior in a parametrically scaled bimanual movement system exhibiting such a bifurcation. In the figure, the displacement-time profiles of left and right hands are plotted against each other on the Lissajous plane. Here the phase relation between the movements of the right and left hands describes the spatiotemporal ordering among corresponding flexor and extensor muscle activities. Starting in the antiphase modal pattern (i.e., right flexion [extension] is accompanied by left extension [flexion]), the parameter of movement frequency is voluntarily increased in a continuous manner. As the frequency increases, the antiphase mode becomes less stable as exemplified by the increase in phase variance. At a critical value (which turns out to be a dimensionless function of each individual's preferred cycling rate), the system bifurcates and a different, in-phase, modal pattern appears (for a more complete analysis, see Kelso, 1984). Extrapolating the above concepts to the postural domain of Nashner and McCollum, we envisage one "discrete strategy" as giving way to another at critical borders in the postural parameter space.

Several points, therefore, are pertinent to Nashner and McCollum's analysis. First, transitions from one synergistic pattern of muscle elements to another may be discontinuous even though the factors controlling the process can change continuously. Second, discontinuities of muscular pattern (giving rise to a description with apparently discrete properties) are observed not because there are no intervening behavioral states, but because none of them is stable (see possible experimental outcomes above). Thus, there may be a large number of ways for a system to exhibit continuous change but only a small number of ways for it to change discontinuously. To conclude, therefore, that discrete logical control is imposed upon a continuous mechanical system may not be warranted. Rather, synergistic muscular activities may emerge as modal patterns from appropriately scaled neuromuscular dynamical systems. Finally, although discrete logical states could be used to represent distinct modal patterns, it should be recognized that much of this apparent discreteness reflects the larger time constants of the dominant modes relative to the time constants of the subsystems. With reference to postural control, the synergistic patterning among muscles appropriate to a given region of the associated parameter space is defined over longer time spans than, say, those involved in motor unit recruitment. Thus, the discrete-logical vs. continuous-dynamical distinction drawn by Nashner and McCollum may be more apparent than real.

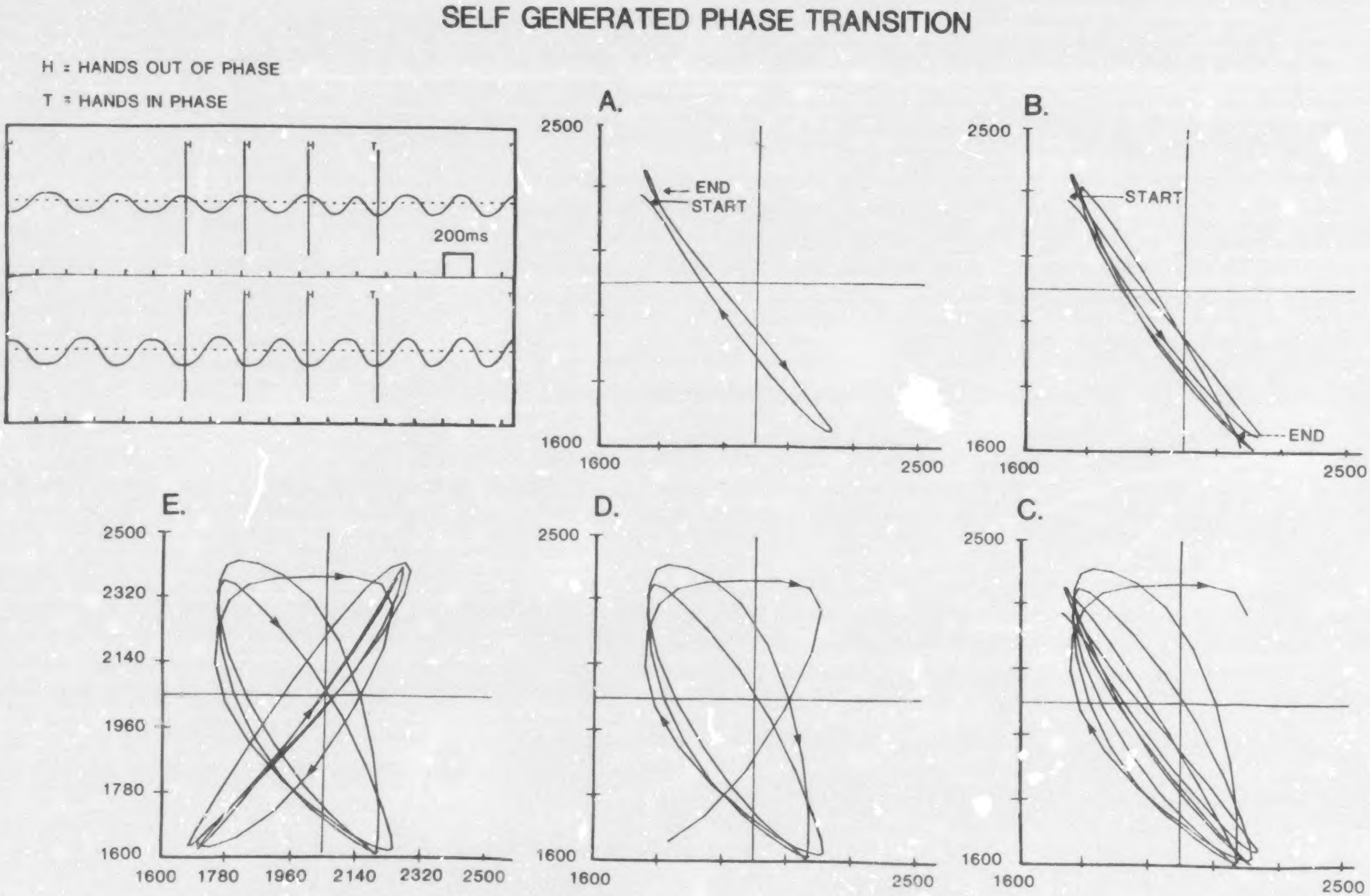


Figure 1. Upper left box shows angular position-time profiles of left (top) and right (bottom) hands. In the remaining plots, the left (x-axis) and right (y-axis) hands are on the Lissajous plane (A-E). "Hands out of phase" means that flexion of one hand is accompanied by extension of the other and vice versa. "Hands in phase" means that both hands flex and extend at about the same time. Phase becomes less stable (1C) as evident in the widening of the Lissajous trajectory, until an abrupt transition occurs (1D). Hand positions on all plots are displayed in arbitrary units (from Kelso & Tuller, 1984).

References

Cohen, A. H., Holmes, P. J., & Rand, R. H. (1982). The nature of the coupling between segmental oscillators of the Lamprey spinal generator for locomotion: A mathematical model. *Journal of Mathematical Biology*, *13*, 345-369.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. *Review of Modern Physics*, *47*, 67-121.

Haken, H. (1977). *Synergetics: An introduction*. Heidelberg: Springer-Verlag.

Haken, H. (1983). *Advanced synergetics: Instability hierarchies of self-organizing systems and devices*. Heidelberg: Springer-Verlag.

Jerard, R. B., & Jacobsen, S. C. (1980). Laboratory evaluation of a unified theory for simultaneous multiple axis artificial arm control. *Journal of Biomechanical Engineering*, *102*, 199-207.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative, and Comparative*, *246*, R1000-R1004.

Kelso, J. A. S., & Saltzman, E. L. (1982). Motor control: Which themes do we orchestrate? *The Behavioral and Brain Sciences*, *5*, 554-557.

Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum Press.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior* (pp. 3-47). New York: Elsevier/North-Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination* (pp. 5-78). Chichester: John Wiley.

Lee, W. A. (1984). Neuromotor synergies as a basis for coordinating intentional action. *Journal of Motor Behavior*, *16*, 135-170.

Litvintsev, A. I. (1972). Vertical posture control mechanisms in man. *Automation and Remote Control*, *33*, 590-600.

Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences*. San Francisco: W. H. Freeman & Co.

Saltzman, E. (1979). Levels of sensorimotor representation. *Journal of Mathematical Psychology*, *20*, 91-163.

Saltzman, E., & Kelso, J. A. S. (1984). Skilled actions: A task dynamic approach. *Haskins Laboratories Status Report on Speech Research*, *SR-76*, 3-50.

von Holst, E. (1973). Relative coordination as a phenomenon and as a method of analysis of central nervous functions (1939). In *The behavioral physiology of animals and man: The collected papers of Erich von Holst* (Vol. 1, pp. 33-135). (R. Martin, Trans.). Coral Gables, FL: University of Miami Press.

Yates, F. E (1982). Outline of a physical theory of physiological systems. *Canadian Journal of Physiology and Pharmacology*, *60*, 217-248.

Yates, F. E., & Iberall, A. S. (1973). Temporal and hierarchical organization in biosystems. In J. Urquart & F. E. Yates (Eds.), *Temporal aspects of therapeutics* (pp. 17-34). New York: Plenum.

REPETITION AND COMPREHENSION OF SPOKEN SENTENCES BY READING-DISABLED CHILDREN[*]

Donald Shankweiler,[†] Suzanne T. Smith,[†] and Virginia A. Mann[††]

Abstract. The language problems of reading-disabled elementary school children are not confined to written language alone. These children often exhibit problems of ordered recall of verbal materials that are equally severe whether the materials are presented in printed or in spoken form. Sentences that pose problems of pronoun reference might be expected to place a special burden on short-term memory because close grammatical relationships obtain between words that are distant from one another. With this logic in mind, third-grade children with specific reading disability and classmates matched for age and IQ were tested on five sentence types, each of which posed a problem in assigning pronoun reference. On one occasion, the children were tested for comprehension of the sentences by a forced-choice picture verification task. On a later occasion they received the same sentences as a repetition test. Good and poor readers differed significantly in immediate recall of the reflexive sentences, but not in comprehension of them as assessed by picture choice. It is suggested that the pictures provided cues that lightened the memory load, a possibility that could explain why the poor readers were not demonstrably inferior in comprehension of the sentences even though they made significantly more errors than the good readers in recalling them.

The problems of many children who are deficient in reading skills are not confined to reading and writing, but extend to abilities involving spoken language as well. Characteristically, the language tasks on which poor readers are deficient place a burden on verbal short-term memory. For example, tasks which require retention of spoken letter names (Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979) word strings and sentences (Mann, Liberman, & Shankweiler, 1980) have consistently distinguished poor readers in the early school

---

[*] In press, Brain and Language.

[†] Also University of Connecticut.

[††] Also Bryn Mawr College.

Acknowledgment. We are grateful to the staff of the East Hartford, Connecticut public schools for their generous cooperation in making this study possible. We wish to record our thanks to the administrative staff, principals, reading consultants, third-grade teachers, and students at the Barnes, Center, Mayberry, Silver Lane, and Stevens Schools. Special thanks are due to Edgar Zurif and Sheila Blumstein for bringing relevant aphasia studies to our attention and for valuable discussion. We also thank Stephen Crain and Carol Fowler for their helpful comments on earlier versions of the paper. The research and preparation of this manuscript was supported by NICHD grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-79/80 (1984)]

years from their peers who are good readers. That the memory problems of the poor readers are language-related is evident from the fact that they typically perform at a level equivalent to good readers on tasks that involve memory for nonlinguistic material such as photographs of faces (Liberman, Mann, Shankweiler, & Werfelman, 1982), visual nonsense designs (Katz, Shankweiler, & Liberman, 1981; Liberman et al., 1982), and visual-spatial sequences (Mann & Liberman, in press).

The purpose of the research we describe here was to investigate the abilities of third-grade children who differ in reading ability to repeat and to comprehend a variety of spoken sentences. Our intent was to explore a possibility that arises from our earlier research (Shankweiler et al., 1979; Mann et al., 1980): that the limitation of verbal short-term memory, which is found to be characteristic of children with reading disability, may be associated with difficulty in spoken sentence comprehension. The expectation that this association would be found was motivated by a consideration of the need for an effective working memory during sentence processing. We assume that a system must exist for holding the words of a sentence and their order of occurrence in some kind of temporary store until the sentence structure can be apprehended. This would follow from the fact that the meaning of a sentence is not merely the sum of the meanings of the separate words it contains, but is derived from the relations between the component words that determine its syntactic and semantic structure. Given poor readers' problems in remembering ordered sequences of words, they might be expected to make mistakes in sentence processing whenever they are confronted with sentences that place the working memory system under stress.

In addition to the sheer number of words a sentence contains, its lexical content and manner of construction can be expected to affect how severely the working memory is taxed in processing it. Sentences with unpredictable or arbitrary semantic content may place a heavy load on working memory because they force the listener to process them fully and perhaps more than once in order to extract the content. The Token Test of De Renzi and Vignolo (1962) contains such structures. This clinical diagnostic test, well-known to students of aphasia, consists of sentence "commands" that request the subject to perform arbitrary manipulations of the token objects. We have found a shortened version of the Token Test (De Renzi & Faglioni, 1978) to distinguish groups of good and poor readers in the third grade, but only on the complex structures in the final sections of the test (Smith, Mann, & Shankweiler, in preparation).

Since most of the Token Test items were insufficiently difficult to separate the good and poor readers, we sought to develop a sentence test that would be at once more sensitive and more analytic. The new measures were designed to discover whether poor readers are selectively impaired in coping with specific types of constructions that stress working memory more by their syntactic form than by their semantic content. Frequently, close grammatical relationships obtain between words that are distant from one another in the string, as in some relative clause sentences in which the logical subject is separated from its pronominal referent by a span of words. Sentences of this form should be very difficult to comprehend if there is inaccurate retention of the word string.

We conducted two additional studies with the same groups of good and poor readers who had received the Token Test. In planning these studies we sought guidance both from the literature on acquisition of syntax by normal children and from studies of sentence comprehension by adults with acquired aphasia. In our first study (Mann, Shankweiler, & Smith, in press) we examined sentences with relative clause structures in which we varied the point of attachment of the relative clause to the main clause. We found that the poor readers made more errors than the good readers on each of four sentence types, but when the four types were ranked in order of difficulty for good and poor readers separately, the ordering was the same for both groups. The finding that the poor readers were generally worse in comprehension of relative clause sentences, but within this broad class, were affected by syntactic variations in the same way that the good readers were, suggests that efficiency of working memory, and not differential grasp of syntactic structure, is the characteristic on which the groups are most readily distinguished.

Thus, the data from studies of sentence memory, the Token Test, and comprehension of relative clause structures are consistent with the possibility that poor readers have deficiencies in sentence processing that are an expression of their difficulties in retaining verbal material in working memory. However, we cannot exclude the possibility that other linguistic deficiencies are present in these children.[1] Although our research to date has not identified any constructions on which poor readers are selectively impaired, we have found that such children usually make more errors in sentence processing than good readers of comparable age and IQ (Mann et al., in press). Poor readers' failures to process sentence materials accurately could reflect memory limitations primarily, as we have suggested, or alternatively, such failures could be symptoms of delayed acquisition of portions of the grammar, as Byrne (1981) has proposed. The possibility that poor readers may have primary syntactic deficits deserves thorough systematic study in which a variety of syntactic structures is examined.

The study we describe here begins to address this need. It focuses on attribution of reference in sentences containing a reflexive pronoun. Our reasons for selecting this problem from among the many possibilities for approaching sentence comprehension were two. First, pronoun reference is tightly governed by syntactic constraints. Since correct attribution of coreference of a reflexive pronoun requires that the perceiver recover the syntactic structure of the whole sentence, comprehension of pronoun reference is a test of sensitivity to grammatical structure. Second, there is evidence that aphasia in adults is often associated with problems in assigning reference to reflexive pronouns. Our study was inspired by an investigation of comprehension of the reflexive by Blumstein, Goodglass, Statlender, and Biber (1983). These investigators compared comprehension of sentences in which a reflexive pronoun is coreferent to an immediately preceding noun phrase, with that of sentences in which the reflexive is coreferent to a noun phrase that occurred earlier in the sentence. Examples 1a and b illustrate these types:

1a The chef watched the boy bandage himself.
1b The chef watching the boy bandaged himself.

Using a two-choice picture-verification task to probe subjects' comprehension of the coreferent of the reflexive in sentences such as 1a and 1b, Blumstein et al. (1983) found that all aphasic subgroups performed better on 1a than on 1b. Indeed, they performed at chance on sentences like 1b, that cannot be successfully comprehended by adherence to a processing strategy in which pronoun reference is inflexibly attributed to the nearest preceding noun phrase. Thus, Blumstein et al. (1983) concluded that the aphasic subjects failed to process fully the syntactic structure of sentences like 1a and 1b, and that they apparently had a tendency to revert to the immature "minimum distance" strategy often attributed to young children (Chomsky, 1969).

Further motivation for our decision to examine children's comprehension of constructions containing reflexive pronouns came from studies that specifically examined developmental changes in pronoun comprehension. Solan (1981) has shown that children of age five or younger recognize the basic constraints on reflexive pronouns. It must be acknowledged, however, that young children do make mistakes in processing pronouns. We note in this connection findings of Read and Hare (1979), who suggest that certain nuances of pronoun use, which turn on the correct parsing of sentences involving more than one clause, may be late to mature. Among a group of children aged six to twelve studied by these investigators, only the oldest subjects in the sample gave grammatically correct interpretations to all types of multiclause constructions that incorporated reflexive pronouns, and even the most successful were not as consistent as adult subjects. Thus, although children may very early apprehend constraints on pronoun reference, considerable individual variation in sophistication in handling reflexive pronouns in multiclause structures seems to exist, giving ample scope for differences between good and poor readers at the third-grade level.

Attribution of pronoun reference seemed, then, to be an important area for further investigation. Accordingly, our study was designed to assess comprehension and immediate recall of sentences containing pronouns. Third-grade children who were good and poor readers were first tested for sentence comprehension by a picture verification test; in a subsequent session on a different day the same sentences were presented for immediate recall.

Method

Subjects

The subjects were 35 third-grade children attending the public school system of a small Northeastern city. All were native speakers of English with no known speech or hearing deficiencies, who had an intelligence quotient of 90 or better, as measured by the Peabody Picture Vocabulary Test (Dunn, 1965). Their inclusion in the experiment was initially based on teachers' evaluations of reading ability, and confirmed by scores on the reading subtest of the Iowa Test of Basic Skills (Hieronymus & Lindquist, 1978), which had been administered approximately four months before our study. Three boys and fifteen girls whose mean Iowa grade-equivalent score was 4.59 (range = 4.1 to 5.2) comprised the good reader group; nine boys and eight girls whose mean Iowa grade-equivalent score was 2.32 (range = 1.7 to 2.6) comprised the poor reader group.[2] The groups did not differ significantly in IQ (109.3 for good readers and 107.7 for poor readers), nor in age (110.5 months for good readers; 107.4 months for poor readers).

Materials

The test materials (see Appendix) consisted of eight tokens of each of five sentence types: Each sentence poses a problem in perception of pronoun reference. A sample set appears below:

A) The fireman watched the soldier bandage himself.

B) The fireman watching the soldier bandaged himself.

C) The fireman bandaged her.

D) The soldier bandaged himself.

E) The soldier bandaged him.

Type A sentences are declarative sentences in which the reflexive pronoun occurs in a relative clause modifying the object of the main clause, thus causing the referent of the reflexive to be the object of the main clause. The pronoun reference can be correctly assigned following the minimum distance principle, since the pronominal referent is the agent immediately preceding the reflexive pronoun. Type B sentences are declarative sentences with a single, center-embedded, relative clause that modifies the subject of the main clause, thus causing the referent of the reflexive pronoun to be the subject of the main clause. In contrast to Type A sentences, the referent of the pronoun in type B sentences cannot be correctly assigned by following the minimum distance strategy, since it is the agent most remote from the reflexive.

The remaining three types of sentences were controls designed to assess comprehension of personal and reflexive pronouns in single-clause sentences. Type C sentences tested the comprehension of personal pronouns, incorporating gender difference as a cue for establishing reference. Types D and E tested comprehension of reflexive and personal pronouns, respectively, without the gender cue.

Eight sentences of each type were constructed using noun agents that can be unequivocally represented and verbs that refer to actions that can be illustrated clearly in drawings. Half of the sentence sets employed male agents and half employed female agents, with Type C sentences incorporating agents of different sexes. The 40 test sentences were randomized and recorded by a speaker who read each one aloud with natural intonation. Each sentence was preceded by an alerting stimulus (a bell).

The tape for the repetition task was recorded separately. It included the original sentences of the comprehension test interspersed with an additional eight control sentences. These control sentences equalled or slightly exceeded the length of Type A and B sentences and incorporated the same agents and actions, but lacked reflexive pronouns. Each was of the form "The nurse and the policewoman sprayed water on the flowers." (see Appendix).

Picture-verification test: In order to assess the ability of subjects to comprehend the reflexive pronoun in each type of construction, we created a four-alternative, forced-choice picture verification task in which subjects were presented with a two-by-two array of line drawings and were asked to point to the drawing that most accurately depicted the meaning of the sentence as heard. The response array for each sentence included four 5 x 3 3/4 inch pictures, one correctly depicting sentence meaning, and three foils, each depicting an incorrect interpretation of the sentence. Each picture displayed two agents; the placement of the agents remained constant within an array, and was varied randomly across arrays. The position of the correct picture and the three different foils was varied so that each appeared with equal frequency in each of the four possible positions within the array.

The foils for sentence Types A and B provided the critical measures. Foil 1 for Type A sentences depicted the reflexive pronoun contained in the subordinate clause as incorrectly attributed to the subject of the main clause. Foil 1 for Type B sentences correctly depicted the actions expressed by each verb, but depicted the reflexive as incorrectly attributed to the object of the subordinate clause. This foil provided the test of whether subjects were following a minimum distance strategy, an assignment that was characteristic of adult aphasics studied by Blumstein et al. (1983). Foil 2 for both Type A and B sentences allowed a test of whether the subject had attended to the entire sentence. This foil depicted the correct attribution of the reflexive to its referent, but incompletely represented the relation between the agents indicated by the first verb. For example, in sentence A (see above), the nurse is not watching the policewoman, and in B, the policewoman is not watching the nurse. Foil 3 for A and B sentences allowed the reflexive pronoun to be interpreted as a personal pronoun.

Foils for the control sentences (C, D, and E) were as follows: Foil 1 depicted reversed roles of the two noun agents. Foil 2 depicted the pronoun incorrectly--i.e., personal pronouns in Type C and E sentences were depicted as reflexive pronouns; reflexive pronouns in Type D sentences were pictured as personal pronouns. Foil 3 depicted a role reversal and misrepresented the pronoun as described above.

Procedure

Subjects were tested individually in two half-hour sessions. The comprehension test was administered first followed by the repetition test at least one week later. When testing comprehension, the examiner placed the relevant array of pictures before the subject immediately prior to the initiation of each tape-recorded sentence. The decision to expose the picture array before sentence onset was dictated by a concern not to overload short-term memory. Subjects were instructed to listen to the whole sentence, to examine each of the four pictures, and then to point to the one that best showed what the sentence meant. Emphasis was placed upon listening to the entire sentence before pointing, and choosing the picture only after examining all of the alternatives. A bell signalled the onset of each test sentence. If a subject requested that a sentence be repeated, the experimenter replayed the sentence once, noting the repetition on the score sheet.

In the sentence repetition task, subjects were instructed to listen to each taped sentence and to repeat it back immediately. Each sentence was played only a single time. If a child requested that a sentence be repeated, the examiner encouraged him to report as much as could be remembered. The responses were transcribed by the experimenter during the session, and also preserved on tape for later error analysis.

Results

Sentence Repetition

The repetition data were analyzed both in terms of the number of incorrectly recalled sentences, and in terms of the total number of individual errors made, including omissions, substitutions, reversals, tense changes, and pronoun errors within each sentence. The results of each scoring procedure are summarized in Table 1 for each type of sentence (the five test types A-E and the additional control type), separately for good and poor readers.

Table 1

Sentence Repetition: Mean number of sentences incorrectly recalled (max=8) and mean number of words incorrectly recalled in sentences of each type

Sentence Type	Reader Group Good (N=18) Mean (SD)	Reader Group Poor (N=17) Mean (SD)
Sentences		
A	2.22 (1.55)	3.41 (1.70)
B	2.06 (2.01)	3.82 (2.19)
C	0.39 (0.92)	1.00 (0.79)
D	0.22 (0.55)	1.23 (1.09)
E	0.11 (0.32)	0.88 (0.99)
Control	2.00 (1.68)	2.70 (1.83)
Words		
A	3.06 (2.31)	4.94 (2.33)
B	3.89 (5.26)	7.35 (7.44)
C	0.39 (0.98)	1.00 (0.79)
D	0.22 (0.55)	1.41 (1.28)
E	0.11 (0.32)	1.06 (1.25)
Control	3.33 (3.27)	5.47 (4.39)

Poor readers made more errors than good readers on both the number of sentences and the number of words to be recalled. Pearson product-moment correlation coefficients were computed for each error measure and the reading scores from the Iowa test. Each was negatively correlated with reading ability: $r(35) = -.48$, $p < .01$ for sentences; $r(35) = -.45$, $p < .01$ for words. Each set of error measures was also subjected to an analysis of variance in which type of sentence (Types A-E and the control sentences) was the within-subjects factor and reading group the between-subjects factor. Significant main effects were obtained for type of sentence, both for number of sentences incorrectly recalled, $F(5,165) = 37.81$, $p < .001$ and number of words, $F(5,165) = 21.97$, $p < .001$. The effects of reader group were also significant: $F(1,33) = 8.80$, $p < .006$ for sentences, $F(1,33) = 6.40$, $p < .017$, for words. However, there was no interaction between reading ability and the ef-

Table 2

Distribution of repetition errors according to word class and error type

(Mean number errors per subject)

	ERROR TYPE									
	Substitution		Deletion		Intrusion		Inflection[2]		PERCENT OF TOTAL ERRORS	
READER GROUP:[1]	Good	Poor	Good	Poor	Good	Poor	Good	Poor	Good	Poor
WORD CLASS										
Nouns	3.44	6.06	0.55	1.23	0.00	0.06	0.39	0.12	39.8	35.2
Verbs	0.83	1.17	0.28	0.53	0.00	0.12	1.89	3.47	27.3	24.9
Pronouns	1.00	3.76	0.55	0.41	0.17	0.29	NA	NA	15.6	21.0
Article	1.00	1.47	0.55	1.53	0.00	0.00	NA	NA	14.1	14.2
Prep./Conj.	0.11	0.35	0.17	0.23	0.06	0.41	NA	NA	3.1	4.7
PERCENT TOTAL ERRORS:	58.0	60.4	19.1	18.5	2.1	4.2	20.7	16.9		

[1] Good readers: N=18; poor readers: N=17

[2] Not applicable for articles, pronouns, prepositions, and conjunctions

fect of sentence type. For children in both reading groups, more errors were made on Type A and B sentences and length-matched control sentences, than on Types C, D, and E, $t(33) = 6.87$, $p < .001$.

Table 2 displays the distribution of errors for each reader group according to error type and word class. The greatest proportion of errors for both reader groups occurred on nouns and verbs. Substitutions within word class, e.g., saying a "a" for "the," "fireman" for "farmer," "hisself" for "himself," make up the greatest proportion of errors for both reader groups. The proportion of deletion errors (deletion of whole words) and errors involving inflections (e.g., omission of the possessive "s"; omission or change of verb tense markers) was comparable for each group. Intrusions, i.e., inserting extra words into a sentence, occurred rarely. It is apparent from Table 2 that although the poor readers made more errors than the good readers in most error categories, the distribution of the errors is highly similar in the two groups.

Sentence Comprehension

Having established that the poor readers were less accurate in verbatim repetition of the test sentences, we turned next to the results of the measure of sentence comprehension, the four-choice picture verification test. The initial analysis was performed on the number of error responses made on each sentence type (A-E). The correlation between total errors and the Iowa score yielded a nonsignificant value of $r(35) = -.14$. Analysis of variance for the factors sentence type and reader group revealed a highly significant effect of sentence type, $F(4,132)=38.06$, $p < .001$, but no significant difference between children in the two reading groups, $F(1,33)=0.40$. Moreover, there was no interaction between individual sentence type and reader group, $F(4,132)=1.53$.

Table 3 shows a breakdown of the errors by sentence t[illegible] and serves to confirm the absence of interaction between the reading groups. It may be seen that many more errors occurred on sentences A, B, and E, than on C and D. The difference between A and B on the one hand, and C and D, on the other, was expected. The comparatively high error rate on Type E may have occurred for a special reason.[3]

A detailed analysis of the error pattern was undertaken in which choice of foils was examined for the critical Sentence types A and B, which were designed to indicate whether poor readers tend to adopt a minimum distance strategy in assigning a referent to the reflexive pronoun. An analysis of variance was performed on this portion of the error data, in which the factors were sentence type, foil type, and reader group. There was a significant effect of sentence type, $F(1,33)=31.53$, $p < .001$, and foil type, $F(2,66)=4.64$, $p < .02$. Moreover, there was an interaction of foil type and reading ability, $F(2,66)=4.02$, $p < .03$. However, there was no interaction of foil type x sentence type x reading ability.

The distribution of errors across the foils for Type A and B sentences is shown in Table 4. The figures in this table are a breakdown of the error means shown in Table 3 according to foil type. Foil 1 in Type B sentences provided the critical test of adherence to the minimum distance principle. Choice of this foil would indicate that in the assignment of pronoun reference, the subject is using a minimum distance strategy in lieu of full syntactic analysis. This was the error that aphasic patients, studied by

Table 3

Sentence Comprehension: Mean number and percent of errors on sentences of each type (max=8)

Sentence Type	Reader Group: Good (N=18) Mean (SD)	Good (N=18) Percent	Poor (N=17) Mean (SD)	Poor (N=17) Percent
A	1.56 (0.92)	18.35	1.35 (1.00)	14.27
B	3.11 (2.08)	36.59	3.88 (2.44)	41.01
C	0.78 (0.65)	9.18	0.35 (0.70)	3.70
D	0.50 (0.71)	5.88	0.59 (0.71)	6.23
E	2.55 (2.12)	30.00	3.29 (1.79)	34.79

Table 4

Distribution of Errors by Foil Type for Sentence Types A and B: Mean number errors

	Foil Type	Reader Group: Good (N=18) Mean (SD)	Poor (N=17) Mean (SD)
Sentence Type A	1	0.00 (0.00)	0.18 (0.39)
	2	0.56 (0.70)	0.29 (0.84)
	3	1.00 (0.68)	0.88 (0.78)
Sentence Type B	1	1.50 (1.85)	2.82 (2.76)
	2	0.78 (0.73)	0.41 (0.62)
	3	0.83 (0.99)	0.65 (0.70)

Blumstein et al. (1983), tended to make. The subjects of the present study also showed a tendency to make this error, that is, they tended to assign the reference to the agent in closest proximity rather than to the referent dictated by the syntax. However, although the poor readers selected Foil 1 more frequently than good readers, the difference was not confined to Type B sentences, as indicated by the lack of a three-way interaction among sentence type, foil type, and reader group. The poor readers tended instead to make more errors on Type 1 foils for all sentence types, $t(33)=1.92$, $p < .05$, suggesting that their difficulty cannot be understood as an inordinate reliance on the minimum distance strategy. Had this been the case, the poor readers should not have made more Foil 1 errors than the good readers on Type A sentences in which Foil 1--in violation of the minimum distance principle--incorrectly attributed the reflexive to the subject of the main clause. As for the other foils, any differences between good and poor readers failed to reach significance. Selection of Foil 2, which controlled for inattention to the first verb of the sentence in both Type A and B sentences, occurred only rarely in either sentence type. Foil 3, which depicted the reflexive pronoun as a personal pronoun in both sentence types, was selected slightly more frequently, but differences between reader groups were minimal.

Selection of foils on the control sentences (C, D, and E) also showed no reader group differences. The few errors that occurred on Type C and D sentences, involved primarily Foil 2, that is, treating a personal pronoun as a reflexive, or vice versa. As we mentioned earlier, somewhat more errors occurred on Type E sentences. These errors predominantly involved personal pronouns in locative constructions (Sets 4 and 6 in Appendix) and indirect object constructions (Sets 2, 5, and 8 in Appendix) having been misinterpreted as reflexive pronouns (choice of Foil 2). Such misinterpretations are common to many young children and may reflect a tendency to "flatten" embedded structures (Tavakolian, 1981).

Discussion

This study was undertaken as part of a continuing investigation of the nature of language impairment in children who fail to make expected progress in learning to read. Here we have asked whether poor readers' problems with language extend to the processing of multiclause spoken sentences involving attribution of pronoun reference. To this end we have tested good and poor readers' repetition and comprehension of the same set of sentences.

With respect to repetition, more errors occurred on the longer, complex sentences. Structural differences between sentences matched for length were not significantly reflected in error rates, although fewer errors tended to occur on sentences that could be interpreted by following the minimum distance principle (Type A). The poor readers overall were less accurate than the good readers in repeating sentences of every type. Sentence type did not significantly affect the extent of differences related to reading ability when the data are examined for number of correct responses and for the pattern of errors. This is in keeping with a finding we reported earlier (Mann et al., 1980) in which it was demonstrated that good and poor readers, similar to the present subjects, though a year younger, differed markedly in recall of both meaningful and meaningless sentences, but the differences were constant across a variety of sentence structures. The results of both studies are consistent with the many lines of evidence that implicate working memory in the language-related deficits of poor readers.

The test for comprehension of the sentences by the picture verification task revealed appreciably more errors on complex sentences than on simple ones. The errors were confined chiefly to multiclause constructions and to the specific locative and indirect object structures that have been identified by other investigators as sources of potential confusion in young children (e.g., Read & Hare, 1979; Roeper, 1982; Solan, 1981). The comparison of greatest interest, between sentences that can be interpreted by following the minimum distance principle (Type A) and those that cannot (Type B), revealed that significantly more errors occurred on the latter, suggesting that the children in our study resorted occasionally to immature parsing strategies. Unlike the repetition test, however, the picture verification test of comprehension did not significantly distinguish the good and poor readers. Such difficulties as the subjects did encounter were common to both groups of children. The children's difficulties with the more complex structures were minor in comparison to the problems that the aphasic patients of Blumstein et al. (1983) encountered with similar sentences. The aphasics performed at chance level on all sentences in which the structure did not allow application of the minimum distance principle, and, indeed, they failed to interpret reflexive pronouns correctly even in simple sentences.

Though these results did not reveal the expected differences between the good and poor readers in comprehension of complex sentences containing reflexive pronouns, we must acknowledge, and take account of, other indications that our good and poor readers are not wholly equivalent in their abilities to comprehend spoken sentences. First, we should note that the children in our two reading groups did not perform equivalently on the reading subtest of the Iowa Test of Basic Skills. The inferior performance of the poor reader group on this test of reading comprehension does not necessarily indicate language processing limitations as such; it may instead reflect limitations that are specific to written language, such as slow and inaccurate word decoding. By studying comprehension of spoken sentences, we hoped to gain a perspective on possible language comprehension limitations, independent of specific reading difficulties. In this connection, it is appropriate to refer to a companion study to the present one in which we tested the same groups of subjects on a different occasion with a different set of sentences (Mann, Shankweiler, & Smith, in press). In that study, unlike the present study, the poor readers displayed a significant deficit in comprehension. There, the method of testing was by object manipulation, not picture verification. Thus the answer to the question of whether the poor readers are below par on comprehension may depend on which structures are assessed and on the method of testing.

Little information is presently available about the capabilities of good and poor readers to comprehend various types of sentences. A recent study by Byrne (1981), which came to our attention after this experiment and the one of Mann et al. were completed, also finds differences in sentence comprehension (as tested by object manipulation) on some sentence types but not on others. The sentences that separated the reader groups in Byrne's study contained unusual constructions and semantic anomalies. Having found that some shorter sentences distinguished the reader groups more readily than longer ones, Byrne argued that memory factors could not be responsible for the differences. This conclusion does not necessarily follow. As we noted earlier, more is involved in memory-related difficulty than sentence length alone. Anomalous sentences, even if they are short, may place extra-heavy demands on working memory because they are likely to be misinterpreted on first construal and therefore need to be "replayed" from memory, in order to establish their structure properly. Such rehearsal would require complete retention.

In regard to the method of testing, we may speculate that the picture verification task of the present study may have stressed short-term memory less than the "acting out" manipulation task of Mann et al. (in press). It is pertinent that in the present experiment, the subjects were allowed to inspect the sheet containing the four multiple-choice picture foils as the sentence was being read, a procedure that could be expected to minimize the need for rehearsal. In contrast, the manipulation procedure of the Mann et al. study merely presented the child with a random arrangement of the three relevant actors (toy animals) in advance of presentation of the sentence. It is clear that the picture test gives more concurrent information, and thus might be expected to stress working memory significantly less. This speculation is supported by the findings of Elmore-Nicholas and Brookshire (1981), in which performance of aphasic adults on a sentence verification task was facilitated by the presence of pictures. Thus, there may be no real inconsistency in the findings of the two studies that tested sentence comprehension in these subjects. Conceivably, the present experiment failed to detect real differences between the reading groups because the method of testing did not give adequate scope for differential performance.

In summary, the poor readers of this investigation were less accurate than the good readers in immediate recall of sentences containing reflexive pronouns, but were not deficient in comprehension of the same sentences. They were deficient, however, both in recall and interpretation of another set of complex sentences, as reported by Mann et al. (in press). We suspect that the comprehension testing conducted with these children yielded inconsistent results because the picture verification procedure used to test reflexive pronouns was insufficiently sensitive. The performances of the poor readers did not closely resemble those of the adult aphasics studied by Blumstein et al. (1983). Unlike the aphasics, neither good nor poor readers displayed rigid adherence to a minimum distance strategy for determining pronoun reference. Nevertheless, reading disabled children--the present group included--have not typically been found to be the equals of good readers in processing spoken sentences (Byrne, 1981; Mann et al., in press), nor, as we have noted, in the use of short-term memory codes which so often are impaired in aphasia (Goodglass, Denes, & Calderon, 1974; Martin & Caramazza, 1982). It seems important, therefore, to explore fully the relations between short-term memory deficits and sentence-processing deficits, and in this regard to seek a better understanding of the similarities and differences between developmental language disorders, such as specific reading disability, and linguistic deficits in the acquired aphasias.

References

Blumstein, S., Goodglass, H., Statlender, S., & Biber, C. (1983). Comprehension strategies determining reference in aphasia: A study of reflexivization. Brain and Language, 18, 115-127.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? Applied Psycholinguistics, 2, 201-212.

Chomsky, C. (1969). The acquisition of syntax in children from five to ten. Cambridge, MA: MIT Press.

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. Cortex, 14, 41-49.

De Renzi, E., & Vignolo, L. A. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasia. *Brain*, *85*, 665-678.

Dunn, L. M. (1965). *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Elmore-Nicholas, B., & Brookshire, R. H. (1981). Effects of pictures and picturability on sentence verification by aphasic and nonaphasic subjects. *Journal of Speech and Hearing Research*, *24*, 292-298.

Goodglass, H., Denes, G., & Calderon, M. (1974). The absence of correct verbal mediation in aphasia. *Cortex*, *10*, 264-269.

Hieronymus, A. N., & Lindquist, E. F. (1978). *Iowa Test of Basic Skills*. Boston: Houghton Mifflin Co.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology*, *32*, 474-484.

Liberman, I. Y., & Mann, V. A. (1981). Should reading remediation vary with the sex of the child? In A. Ansara, N. Geschwind, A. Galaburda, M. Albert, & N. Gartrell (Eds.), *Sex differences in dyslexia*. Baltimore, MD: The Orton Society.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. *Cortex*, *18*, 367-375.

Mann, V. A., & Liberman, I. Y. (in press). Phonological awareness and verbal short-term memory: Can they presage early reading problems? *Journal of Learning Disabilities*.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory & Cognition*, *8*, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (in press). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language*.

Martin, R. C., & Caramazza, A. (1982). Short-term memory performance in the absence of phonological coding. *Brain and Cognition*, *1*, 50-70.

Read, W. C., & Hare, V. C. (1979). Children's interpretations of reflexive pronouns in English. In F. R. Eckman & A. J. Hastings (Eds.), *Studies in first and second language acquisition*. Rowley, MA: Newbury House.

Roeper, T. (1982). On the importance of syntax and the logical use of evidence in language acquisition. In S. A. Koczaj II (Ed.), *Language development* (Vol. 1, *Syntax and semantics*). Hillsdale, NJ: Erlbaum.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 531-545.

Solan, L. (1981). The acquisition of structural restrictions in anaphora. In S. L. Tavakolian (Ed.), *Language acquisition and linguistic theory*. Cambridge, MA: MIT Press.

Smith, S. T., Mann, V. A., & Shankweiler, D. (in preparation). *Spoken sentence comprehension in good and poor readers: A study with the Token Test*. Unpublished manuscript, Haskins Laboratories.

Tavakolian, S. L. (1981). The conjoined-clause analysis of relative clauses. In S. L. Tavakolian (Ed.), *Language acquisition and linguistic theory*. Cambridge, MA: MIT Press.

Wolford, G., & Fowler, C. A. (1984). Differential use of partial information by good and poor readers. *Developmental Review*, *4*, 16-35.

Footnotes

[1] Nor can we exclude the possibility that the strategies they employ on certain other cognitive tests may be deviant (see Wolford & Fowler, 1984).

[2] The groups were thus not equivalent in the proportion of boys and girls. We do not regard this as a serious imbalance, however, since research has shown that the patterns of deficits characteristic of children with reading disability do not vary with the sex of the child (Liberman & Mann, 1981).

[3] The higher error rate on Type E sentences than on Types C and D, which were matched with these for length, requires comment. E sentences were designed as controls to test basic grasp of pronoun use, and therefore few errors were anticipated from children in the age range of our subjects. The analysis revealed that the principal error on this sentence type was to interpret a pronoun as though it were a reflexive. Thus, the sentence "The astronaut poured him a drink" was interpreted to mean that the astronaut poured a drink for himself. We speculate that this interpretation reflects a dialect preference and not a genuine confusion in assigning pronoun reference. In support of this, we note that on Type C sentences, where reference is established by gender, such misinterpretations practically never occurred.

Appendix

Sentences used in comprehension and repetition

Set

I.A. The fireman watched the soldier bandage himself.
B. The fireman watching the soldier bandaged himself.
C. The fireman bandaged her.
D. The soldier bandaged himself.
E. The soldier bandaged him.
II.A. The astronaut watched the sailor pour himself a drink.
B. The sailor watching the astronaut poured himself a drink.
C. The sailor poured her a drink.
D. The astronaut poured himself a drink.
E. The astronaut poured him a drink.
III.A. The farmer watched the Indian pull himself up the rope.
B. The farmer watching the Indian pulled himself up the rope.
C. The policewoman pulled him up the rope.
D. The Indian pulled himself up the rope.
E. The farmer pulled him up the rope.
IV.A. The clown watched the boy spill paint on himself.
B. The boy watching the clown spilled paint on himself.
C. The girl spilled paint on him.
D. The clown spilled paint on himself.
E. The boy spilled paint on him.
V.A. The girl watched the grandmother make herself a sandwich.
B. The girl watching the grandmother made herself a sandwich.
C. The Indian made her a sandwich.
D. The grandmother made herself a sandwich.
E. The girl made her a sandwich.

VI.A. The nurse watched the policewoman spray perfume on herself.
B. The policewoman watching the nurse sprayed perfume on herself.
C. The clown sprayed perfume on her.
D. The nurse sprayed perfume on herself.
E. The nurse sprayed perfume on her.
VII.A. The waitress watched the ballerina dress herself.
B. The waitress watching the ballerina dressed herself.
C. The nurse dressed him.
D. The ballerina dressed herself.
E. The waitress dressed her.
VIII.A. The witch watched the queen pick herself a flower.
B. The queen watching the witch picked herself a flower.
C. The queen picked him a flower.
D. The witch picked herself a flower.
E. The queen picked her a flower.

Control Sentences (repetition)

1. The sailor and the fireman poured coffee from the pot.
2. The astronaut and the sailor bandaged the boy's hand.
3. The boy and the Indian pulled the sled up the hill.
4. The clown and the farmer spilled paint on the sidewalk.
5. The queen and the grandmother made sandwiches for lunch.
6. The nurse and the policewoman sprayed water on the flowers.
7. The witch and the ballerina dressed for the party.
8. The waitress and the girl picked flowers in the park.

SPELLING PROFICIENCY AND SENSITIVITY TO WORD STRUCTURE*

F. William Fischer,† Donald Shankweiler,†† and Isabelle Y. Liberman††

Abstract. The connection between spelling and pronunciation in many English words is somewhat remote. To spell accurately, a writer may need to appreciate that the orthography maps regularities of more than one kind. Two experiments explored the possibility that young adults who differ in spelling ability also differ in sensitivity to morphophonemic structure and word formational principles that underlie the regularities of English spelling. In the first, an analysis of misspellings showed that poor spellers were less able than good spellers to exploit regularities at the surface phonetic level and were less able to access the underlying morphophonemic structure of words. A second experiment used pseudowords to extend these findings and to confirm that spelling competence involves apprehension of generalizations that can be applied to new instances.

All would agree that English spelling is not easily mastered. Even accomplished readers and writers may at times be uncertain about the spelling of particular words. There is less agreement about why English causes so much difficulty. The reason most often given for spelling failures is the supposed irregularity of English orthography. This diagnosis, though popularly accepted, is a misleading oversimplification. It reflects the widespread confusion about how the orthography represents word structure. An example will serve to illustrate that when English spelling departs from one-to-one correspondence with pronunciation, as it so often does, it may nevertheless preserve orderliness at some other level. The plural _s_ in _cats_ receives an s-sound while the _s_ in _dogs_ is pronounced as _z_. We do not balk at this inconsistency perhaps because the convenience of representing the plural morphophoneme in a consistent way overrides considerations of strict one-to-one correspondence with pronunciation.

It is characteristic of English that the degree of transparency of the mapping between word components and their orthographic representation varies

---

*_Journal of Memory and Language_, in press.

†Now at Central Connecticut State University.

††Also University of Connecticut.

Acknowledgment. This report is based on a doctoral dissertation, _Spelling Proficiency and Sensitivity to Linguistic Structure_, presented by the senior author to the University of Connecticut (Fischer, 1980). We are indebted to several colleagues for suggestions and insightful criticisms of earlier versions of the paper: Vicki Hanson, Leonard Katz, Alvin Liberman, Ignatius Mattingly, Bruce Pennington, and Michael Studdert-Kennedy. The research and the preparation of the manuscript were supported in part by a grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994).

considerably from word to word. This diversity is a consequence of the many and varied sources of the English vocabulary. There are, on the one hand, words like *harp*, which have a morphophonemic structure, and hence a spelling, that is in close correspondence to a typical phonetic realization of the word. On the other hand, there are words in which the morphophonemic structure for one or more segments is at some remove from a phonetic realization of the word. This occurs frequently in words that are foreign borrowings (for example, *bourgeois*) or in words reflecting archaic forms (for example, *gnaw*). In contrasting these two extremes we might characterize the mapping for the first set of words as being all but transparent, whereas the mapping of the second set is relatively opaque to many, perhaps most, users of English.

Many English words have a degree of orthographic transparency that lies somewhere between the extremes represented by the examples given above. Many words are more or less straightforward except that they contain a "problem segment." Examples include such words as *thinned*, *misspell*, and *grammar*. At one specific location in each of these the relationship between the morphophonemic and phonetic structure is not immediately transparent in the spelling. In cases such as these, correct spelling could be facilitated by apprehending the morphemic structure (*mis* + *spell* requires retaining both s's), the orthographic conventions (*thin* + *ed* requires doubling the *n*), or the derivational relationships (the identity of the reduced vowel in *grammar* can be uncovered by relating the word to cognate forms in which the same vowel segment is not reduced, as in *grammatical* or *grammarian*).

It is one thing, however, to demonstrate that order exists in the mapping of word and orthography. It is quite another to show that the regularities are apprehended and utilized by ordinary spellers who are not linguistic scholars. If we accept the premise that English orthography is by and large a rational system, it is reasonable to suppose that successful use of the orthography may be dependent on the users' ability to understand the system, or on what we shall call their "linguistic sensitivity."

We use the term "linguistic sensitivity" to refer to the ability to apprehend the inherent regularities at various levels of linguistic representation and the ability to exploit this knowledge in reading and writing words. There exists already considerable evidence that successful readers can be distinguished from unsuccessful ones on a number of metalinguistic abilities (Fowler, Shankweiler, & Liberman, 1979; Liberman, Shankweiler, Fischer, & Carter, 1974; Morais, Cary, Alegria, & Bertelson, 1978; Perfetti & McCutchen, in press; Vellutino, 1979). It is possible that major differences in linguistic sensitivity so defined may also be associated with the large variations in spelling ability that are found even among highly-schooled adults. In the past, investigators have looked repeatedly to nonlinguistic explanations, appealing, for example, to individual differences in visual memory ability (Shaw, 1965; Witherspoon, 1973). The alternative view is that spelling draws heavily upon knowledge of linguistic structure. Although this viewpoint is not new (see, in particular, Chomsky & Halle, 1968), the recent spate of papers on spelling offers little direct empirical evidence either pro or con (but see Frith, 1978; Marcel, 1980; and Steinberg, 1973). The present study was designed to fill what seemed an obvious need.

Before an empirical investigation could be started, test materials capable of assessing sensitivity to the structural properties of the orthography had to be developed. Although some experimental spelling tests (e.g., Barron,

1980) categorize words as "regular" or "irregular," the basis for classification is not usually made explicit. The classification of "regular" is typically applied to words having a presumed straightforward correspondence between spelling patterns and phonetic structure (e.g., fresh). Accordingly, words with regularities of all other kinds are typically designated as "irregular" (e.g., sign), despite their demonstrable adherence to a pattern or rule. A further shortcoming of the available tests is that they are constructed without regard to variations in word frequency. Together, these deficiencies make existing tests unsuitable for our purposes. Accordingly, an Experimental Spelling Test was developed to overcome these limitations. While controlling for word frequency, it attempts to capture some of the structural properties that give rise to different levels of transparency in English spelling.

The hypothesis under investigation is that educated adults who differ in spelling ability on conventional spelling tests differ correspondingly in the knowledge we call linguistic sensitivity. To explore this possibility, two experiments were conducted. In the first, the performance of good and poor spellers was examined using the Experimental Spelling Test. It was anticipated that for all subjects those words in which the morphophonemic representation is at some remove from the phonetic structure would be more often misspelled, other things equal, than those words in which the two levels of representation more nearly coincide. Moreover, if good and poor spellers are primarily distinguished on the basis of their metalinguistic abilities, then the largest differences between the groups on the Experimental Spelling Test ought to occur in spelling the words whose mapping can only be rationalized linguistically. Smaller differences, or no difference, should occur on the opaque words, for the spellings of which the subjects may have to rely chiefly on rote memory.

If college-level adults who differ in spelling proficiency can be distinguished on the basis of their sensitivity to certain structural characteristics of real words, then differences among them should be especially evident on tasks that are free from the effects of word-specific learning. The second experiment of this investigation explored this possibility by comparing the performance of good and poor spellers on tasks that tap certain linguistic abilities presumed to be useful in spelling the words on the Experimental Spelling Test. These abilities include knowledge of abstract spelling patterns, familiarity with principles involving prefixation and suffixation, and ability to use tacit knowledge of English morphophonemics in order to disambiguate reduced vowels. New materials had to be developed for tapping these abilities. Pseudowords rather than actual words were used where necessary to ascertain that the subjects had acquired general principles of orthographic representation that can be applied to new instances.

In addition to the assessment of metalinguistic abilities associated with spelling performance, Experiment 2 also examined the possibility that good and poor spellers may differ in their use of visual retention strategies. Since visual memory is often cited as a major determinant of spelling proficiency (Shaw, 1965; Sloboda, 1980; Tenney, 1980; Witherspoon, 1973), a task assessing visual memory ability for abstract designs was included. It was anticipated that on the linguistic tasks, good spellers would continue to outperform those who were less proficient, while no difference between the groups would emerge on the task of visual memory for designs. Finally, the groups of good and poor spellers were compared on tasks designed to tap broader aspects of literacy, namely, reading skills and vocabulary knowledge.

Experiment 1

The purpose of this experiment was to compare the performance of college-educated adults who differ in spelling proficiency on spelling tasks that incorporate graded changes in orthographic transparency.

Method

Subjects. Two groups of subjects, good spellers (N=18) and poor spellers (N=20), were selected from a larger sample of 88 undergraduate psychology students who responded to a notice inviting them to participate in an investigation of spelling ability. The notice had encouraged people to sign up regardless of their level of spelling proficiency. The 88 initial participants were all native speakers of American English, 21 males and 67 females, ranging in age from 18 to 37 years (mean age=20 years). While they do not constitute a random sample, those participating did represent a broad range of spelling proficiency as indicated by their scores on the spelling section of the Wide Range Achievement Test (Jastak, Bijou, & Jastak, 1965). Grade equivalent scores on the WRAT ranged from 8.4 to 15.7 with a mean of 12.3.

Those identified as good spellers for the purpose of this study performed at or above grade level on the WRAT (mean grade equivalent was 14.4, S.D.=0.51). Those categorized as poor spellers were clearly deficient performing on the average four years below grade level (mean grade equivalent was 10.2, S.D.=0.64). The good speller group included 6 males and 12 females, the poor spellers consisted of 4 males and 16 females.

Stimuli. The chief instrument used was the new three-part Experimental Spelling Test of 120 words. The words were grouped into three levels, 40 in each, differing in the transparency of orthographic representation. For Level 1 words, the phonetic realization is, for any given speaker, reasonably close to the orthographic representation, and the spelling patterns are, for the most part, restricted to those having a high frequency of occurrence in written English. Examples of words so classified are *harp*, *adverb*, and *retort*.

Level 2 words each contain an ambiguous segment involving some departure from straightforward phonetic mapping. They are further partitioned into two subtypes. Level 2A words require either a rote application of established orthographic conventions, or a sensitivity to regularities at the surface phonetic level. For example, a speller may know that the /n/ segment is represented by *nn* in *thinned* but by *n* in *chained*. The experienced writer does this quite mechanically, having learned that in monosyllabic words the final consonant letter is doubled when preceded by a single vowel but not doubled when preceded by a vowel digraph. Indeed, in many instances the graphemic conventions relate to phonetic facts such as those involving lax versus tense vowels. In contrast, Level 2B words draw upon abstract morphophonemic knowledge to derive the spelling patterns for the ambiguous segments. For example, in order to know that the final consonant letter in *confer* is doubled in *conferring* or *conferred* but not in *conference*, a speller must apprehend linguistic regularities relating to stress placement, and how these govern spelling. The generalizations included in the list are described in Appendix 1.

Level 3 words can be derived only partially by using morphophonemic knowledge, since they contain one or more segments that do not generally occur in English or occur with low frequency. Their relative lack of transparency stems from two factors: the words are related to borrowed forms largely obscure to the nonscholar and the nonpolyglot, and their spelling patterns have a much lower frequency of occurrence in English than do the patterns appearing in Level 1 and 2 words. Examples include such words as gnaw, bourgeois, and Fahrenheit.

The three levels were balanced insofar as possible for syllable length (each level approximating a mean of 2.8 syllables) and frequency of occurrence in written English (each level approximating a mean of 6.1 occurrences per 1,014,232 words of natural language text), according to the Kucera and Francis (1967) statistics. Within Level 2 the 2A words had a mean frequency of occurrence of 5.7 versus 6.8 for the 2B words. The 2A words had a mean of 2.4 syllables versus 3.4 for the 2B words. The 120 words (which are listed in Appendix 2) were randomized, and recorded on magnetic tape at 10 s intervals.

Procedure

The subjects were tested in small groups. The testing session lasted for one hour during which the following tasks were administered.

1. Spelling Production Task. The subjects' task was to print each dictated word in the space provided and to attempt every word. Each was repeated once.

2. Spelling Recognition Task. The same items were presented again, this time as a multiple-choice recognition test. The answer sheet offered three alternative spellings for each dictated word and, additionally, a "none of these" option. Each of the three alternatives was phonetically readable as the stimulus word; thus no foil could be eliminated merely on the basis of a gross disparity between the spelling of an item and its phonetic realization. Common misspellings of the stimulus words appeared as foils.

3. Spelling Subtest of the Wide Range Achievement Test. (Jastak et al., 1965). The words from the Level 2 spelling list of the WRAT were recorded on magnetic tape at 10 s intervals. The subjects' task was to print the words in the space provided.

Scoring of Spelling Errors

The following error categories were used to analyze the misspellings:

1. Word Errors were scored for each misspelled word without regard to the number of misspelled segments (for example, when grammar was spelled "grammer" or sergeant as "sargent."

2. Segment Errors were scored for every incorrect spelling pattern, as defined by guidelines established by Hanna, Hanna, Hodges, and Rudorf (1966). Segment errors were further classified as substitutions, omissions, or insertions.

a. Substitution Errors were scored when an incorrect grapheme was used in place of the correct letters. These were further classified as "phonetic substitutions" when the word as spelled captures the word's approximate phonetic shape (as when *rhododendron* was spelled "rododendron" or when *gnaw* was given as "naw") and "nonphonetic substitutions" (for example, when *adverb* was spelled "advert").[1]

b. Omission Errors were scored when a grapheme needed for the orthographic representation of a phonological segment was omitted (for example, *inflate* for "infate").

c. Insertion Errors were scored when an additional grapheme was included (for example, *retort* for "restort").

Results and Discussion

A preliminary step was to establish that the Experimental Spelling Test designed provided a reliable and valid estimate of general spelling ability. A test-retest comparison of word errors carried out on a subset (N=30) of the 88 participants resulted in a reliability coefficient of .97 ($p < .001$) on the Spelling Production Task. The results of a correlational analysis revealed that word error scores on the Spelling Production Task correlated significantly ($r = .84$, $p < .001$) with error scores on a standardized test of spelling achievement, the Wide Range Achievement Test. Together, these results suggest that the test yields a reliable measure of spelling achievement and gives results that are highly comparable to a widely-used conventional test of spelling proficiency.

An analysis of item difficulty on the Spelling Production Task was also conducted to examine for possible floor or ceiling effects. It was found that no word was misspelled by every subject, and even the most difficult words on the list (*desiccate* and *sarsaparilla*) were spelled correctly by at least two of the 88 subjects. Although 20 of the 120 words were never misspelled, no subject obtained a perfect score. The number of misspelled words ranged from 18 to 52 with a mean of 33.9 (S.D. = 8.9).

Spelling Production Task: The locus of spelling difficulty. It is important to discover whether the spelling mistakes made by poor spellers are limited to words having particular orthographic or structural characteristics or whether the difficulties reveal more general deficiencies in transcribing English. To answer this, we first looked at the distribution of misspelled words on the Spelling Production Task across the three levels of orthographic transparency (see Figure 1). The data were analyzed by a two-way analysis of variance in which the between-groups factor was spelling group, the within-groups factor was orthographic level and the dependent variable was the number of word errors. As can be seen in Figure 1, the good and poor spellers differed sharply across each of the three orthographic levels: $F(1,36) = 154.73$, $p < .001$, $MSe = 7.95$, for group; $F(2,72) = 717.44$, $p < .001$, $MSe = 4.57$, for level. The interaction between group and orthographic level was also significant, $F(2,72) = 42.21$, $p < .001$, $MSe = 4.57$. Good spellers made significantly fewer errors at each level than did poor spellers (at Level 1, $t(36) = 4.46$, $p < .001$; at Level 2, $t(36) = 12.64$, $p < .001$; and at Level 3, $t(36) = 7.35$, $p < .001$). It is of interest to note that the interaction remains significant when the group by level analysis is recomputed for Levels 2 and 3 alone, $F(1,36)=13.43$, $p < .001$, $MSe = 27.14$. This suggests that the full interaction effect is not simply a consequence of the greater accuracy of both groups in spelling the orthographically transparent Level 1 words, but instead reflects performance differences all across the range of orthographic transparency.

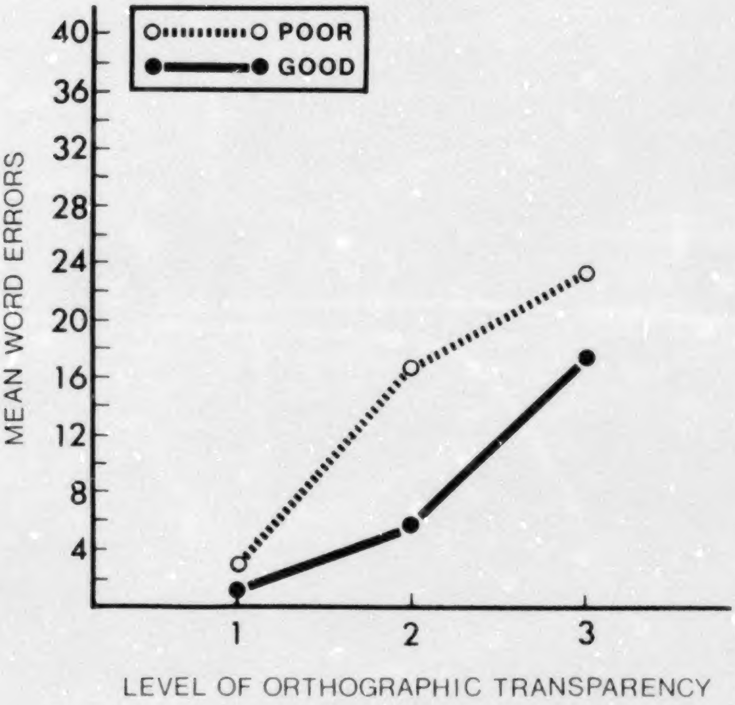


Figure 1. Comparison of word errors on spelling production task as a function of orthographic transparency, good versus poor spellers.

The finding that good and poor spellers differ significantly in their ability to spell words at each of the three levels suggests that they have general deficiencies in spelling rather than isolated, local difficulties restricted to particular exceptional words.

As expected, few Level 1 words were misspelled by either group. Nevertheless, even on these the two groups differed significantly. Errors made by poor spellers were quite varied. In 11 percent of the cases the dictated item was apparently misperceived perhaps because of unfamiliarity with the word--for example, vortex rendered as "thortex" or "vortext." In 29 percent errors occurred in relation to the representation of free versus checked vowels--for example, diplomat rendered as "diplomate", emit as "emite." However, the bulk of the errors (60 percent) were instances of the use of spelling patterns that in another context would be appropriate but are incorrect for the particular morpheme being represented, for example--spelling retort as "rhetort," and punishment as "punnishment." In contrast to the greater range of difficulty experienced by poor spellers, the Level 1 errors of good spellers, with the exception of the word canister (which many spelled "cannister"), were confined to occasional misperceptions of a stimulus word (spelling thinned as "fend" or compensates as "compensate").

Differences in the ability of good and poor spellers to transcribe words are reflected in quantitative differences in virtually every aspect of performance on which the two groups were compared. Table 1 presents an overview of the analysis of segment errors. As anticipated, most errors occurred on those phonologic segments that departed most conspicuously from a straightforward phonetic transcription. As Table 1 reveals for both groups substitution errors accounted for the bulk of the errors made, followed by a much smaller percentage of omissions and even fewer insertions. Overall, the poor spellers made significantly more errors of each type (for substitutions, $t(36) = 8.98$, $p < .001$; for omissions, $t(36) = 3.65$, $p < .001$; and for insertions, $t(36) = 2.42$, $p < .02$). The low percentage of omissions and insertions indicates that both groups were generally accurate in preserving the segmental structure of words.

Table 1

Summary of Segment Errors on Spelling Production Test
Good and Poor Spellers

	Good Spellers			Poor Spellers		
Error Type	Mean	Percent Substitutions	Percent Total Error	Mean	Percent Substitutions	Percent Total Error
Substitutions	31.9	--	85.8	63.2	--	83.9
Phonetic	27.9	87.5	75.0	56.2	88.9	74.6
Nonphonetic	4.0	12.5	10.8	7.2	11.4	9.6
Consonants	12.5	39.2	33.6	22.8	36.1	30.3
Vowels	19.3	60.5	51.9	40.6	64.2	53.9
Omissions	4.4	--	11.8	10.2	--	13.5
Insertions	0.9	---	2.4	1.9	---	2.5
Total Errors	37.2	--	--	75.3	--	--

Since errors of substitution were most numerous, the analysis focused on these. It was found that for both groups significantly more substitutions occurred on vowels than on consonants with the poor spellers again making significantly more errors than good spellers on both consonants ($t(36) = 8.03$, $p < .001$) and vowels ($t(36) = 10.39$, $p < .001$). The greater difficulty in spelling vowel segments is expected since the mapping between orthographic patterns and vowel sounds is generally more variable than it is for consonants. Finally, for both groups phonetic substitutions significantly outnumbered nonphonetic substitutions with the poor spellers again making significantly more of each error type than the good spellers (for phonetic substitutions, $t(36) = 10.88$, $p < .001$; for nonphonetic substitutions, $t(36) = 3.15$, $p < .01$). These data suggest that highly-schooled adults usually represent the phonetic characteristics of words adequately but sometimes fail to attend to the deeper morphophonemic regularities that would have led to the correct spelling.

Production errors versus recognition errors in spelling. In examining the effect of orthographic transparency on spelling accuracy it is of interest to compare the performance of the two groups on the task utilizing a recognition format. Figure 2 presents these data for the good spellers (top) and poor spellers (bottom). The data were analyzed using a three-way analysis of variance in which the between-groups factor is spelling group and the within-groups factors are condition (production and recognition) and level of orthographic transparency (Level 1, 2, and 3). The dependent variable was the number of misspelled words.

As expected, the task of recognizing correctly spelled words proved to be significantly easier for the two groups combined than the task requiring spelling production (for condition, $F(1,36) = 92.32$, $p < .001$, $MSe = 2.68$). The mean word error score under the recognition format was 27.8 compared with a higher mean error score of 34.0 on the production task. No differences were found for the interactions of group by condition, $F(1,36) = 2.54$, $p < .12$, $Mse = 2.68$, or group by condition by level, $F(2,72) = 2.56$, $p < .08$, $MSe = 2.16$. Of particular interest, however, is the finding that for both groups the overall increase in accuracy that occurred under the recognition condition is largely concentrated on the morphophonemically opaque, Level 3 words (for condition by level, $F(2,72) = 97.85$, $p < .001$, $MSe = 2.16$). Whereas subjects typically reduced their word error score on Level 3 words, smaller reductions in errors occurred in spelling the more transparent words. The mean word error score on Level 1 words was 2.0 on the production task versus 1.5 on the recognition task and on Level 2 words, 11.4 mean word errors versus 11.5 mean word errors, respectively.

Differences between good and poor spellers in linguistic sensitivity. While these findings underscore the quantitative differences between good and poor spellers, the critical abilities distinguishing the two groups remain undefined. From a linguistic perspective there are certain skills that still need to be explored. On the one hand, for example, poor spellers might be differentiated from good spellers in their lack of sensitivity to surface orthographic and phonetic regularities that signal the use of particular spelling patterns. Alternatively, or additionally, they might differ in their ability to penetrate below the surface structure to the deeper morphophonemic regularities that determine the appropriate spelling patterns.

In order to evaluate these possibilities, it was useful to examine the performance of good and poor spellers on the Level 2 words where the performance differences between the groups were largest. It will be recalled that each Level 2 word contained an ambiguous segment. In approximately half of the words (Level 2A), the spelling of that segment could be ascertained by recognizing certain orthographic regularities and by implementing the relevant orthographic conventions. In the remaining half (2B), the ambiguous segment could be derived only by accessing the morphophonemic information.

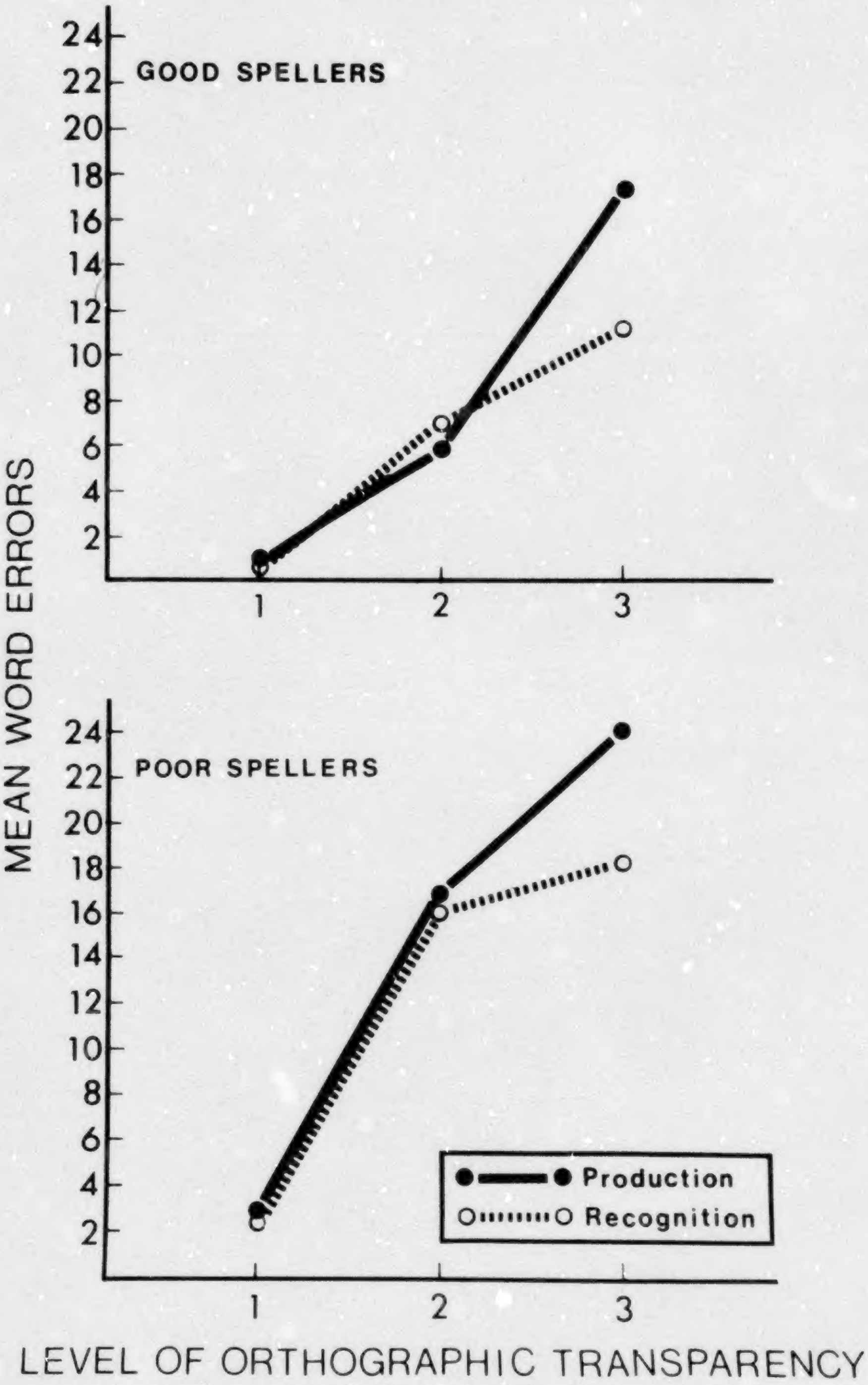


Figure 2. Comparison of word errors on production and recognition tasks as a function of orthographic transparency, good versus poor spellers.

In order to determine whether the good and poor spellers differed in their ability to spell these two subclasses, it was necessary to ascertain whether the errors that occurred did indeed involve the segment designated as the ambiguous segment (the "problem segment"). An examination of the errors revealed that, in both groups, 83 percent occurred on problem segments involving either orthographic or morphophonemic decisions, while the remaining 17 percent occurred on other segments within these words. The analysis was therefore restricted to those errors that occurred at the critical location. In addition, because two spellings were found to be acceptable for one of the Level 2A words (cancelled and canceled, Webster, 1963), it was excluded from the analysis, reducing the total number of words to 19.

In Figure 3 the mean percentage of word errors is presented for Level 2A (orthographic) and Level 2B (morphophonemic) words. The data displayed in Figure 3 were analyzed by a two-way analysis of variance in which the between-groups factor was spelling group and the within-groups factor was error type (orthographic or morphophonemic). The dependent variable was the percentage of word errors based on 19 words in Level 2A and 20 words in Level 2B.

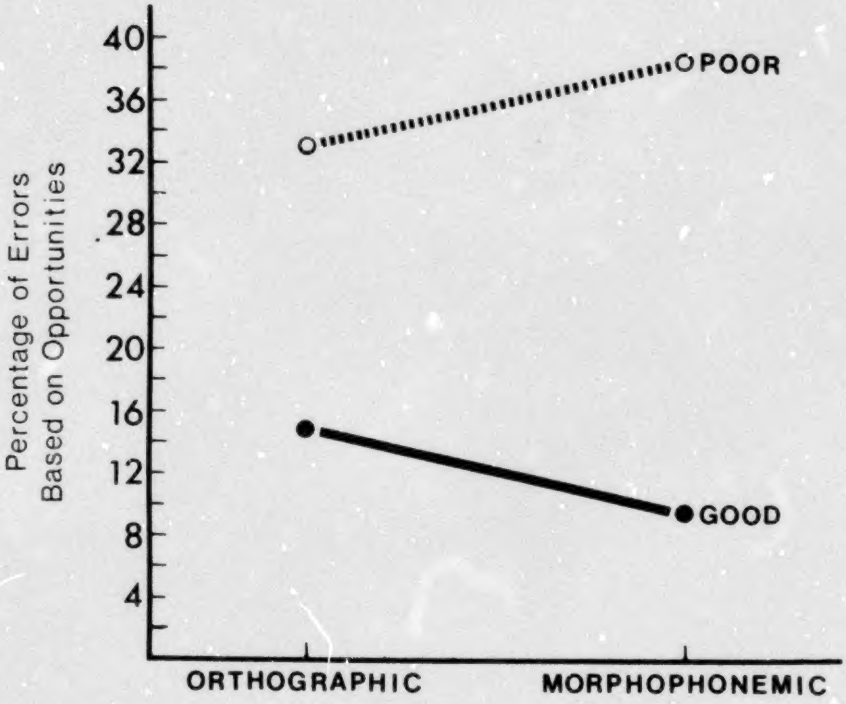


Figure 3. Comparison of orthographic and morphophonemic errors on level 2 words, good versus poor spellers.

Figure 3 shows a wide separation in the performance of the good and poor spellers. Of particular interest, however, is the unequal performance of the two groups on the two categories of words, yielding a significant interaction between group and error type, $F(1,36) = 10.29$, $p < .003$, *MSe* = 51.04. As would be expected, good spellers made fewer errors than poor spellers both in applying orthographic conventions, *Fisher's post hoc* $t(36) = 7.00$, $p < .001$, and in spelling words involving access to morphophonemic structure, $t(36) = 9.54$, $p < .001$. The more notable result, however, is that good spellers found words involving morphophonemic decisions significantly easier than words involving purely orthographic decisions, $t(17) = 2.73$, $p < .02$, while the poor spellers showed no significant difference in their ability to spell the two types of words, $t(19) = 1.98$, $p > .05$. This suggests that good and poor spellers may differ in their ability to penetrate below the surface phonetic structure to the underlying morphophonemic structure of words. To ascertain whether this finding could be generalized to other words not included in the present list, a second ANOVA was computed using the 39 Level 2 words as the random variable (Clark, 1973); the between-groups factor was word type (orthographic vs. morphophonemic) and the within-groups factor was group (good vs. poor spellers). The dependent variable was the percentage of errors made by good and poor spellers on each of the words. The analysis indicated significant effects of word type, $F(1,37) = .02$, $p > .05$; group, $F(1,37) = 60.30$, $p < .001$, *MSe* = 162.21; word by group, $F(1,37) = 6.32$, $p < .001$. As a further step, the *min* F' was computed. The outcome suggests that the differences observed between the groups in spelling the Level 2A and 2B words extend beyond the particular words used in this experiment, *min* $F'(1,54) = 5.0$, $p < .03$.

The contribution of nonlinguistic abilities to spelling proficiency. So far the findings have suggested that differences in spelling achievement are at least in part associated with differences in apprehension of word structure. It is also of interest to examine the results as they relate to a long-held belief that individual differences in spelling proficiency may reflect differences in visual retentiveness. Two aspects of the data are pertinent to this question. If visual memory skill were the critical distinguishing factor, then the greatest performance difference between the groups should occur in spelling the opaque, Level 3 words, since these presumably have to be learned and recalled by rote. However, on re-examining Figure 1, one finds that although good and poor spellers did in fact differ in their ability to spell Level 3 words, the magnitude of the difference is smaller than that which occurred in spelling the derivable, Level 2 words. These results suggest that if there are differences between the groups in their ability to recall visual images of word patterns, these differences are of lesser importance than those relating to the understanding of how the orthography maps word structure.

Moreover, if visual memory ability were an especially critical skill in spelling, good and poor spellers should differ in their ability to recognize correct spellings when given alternatives from which to choose. Reexamination of Figure 2 suggests that the two groups are not readily distinguishable in this regard. This is confirmed by the finding that the relevant interaction effects were not significant (for group by condition, $F(1,36) = 2.54$, $p > .05$, *MSe* = 2.68 or for group by condition by level, $F(2,72) = 2.56$, $p > .05$, *MSe* = 2.16). Thus, on the spelling recognition task good spellers were not significantly better able than poor spellers to profit from visually-presented alternatives. While it is quite likely that visual memory plays some role in spelling (especially for Level 3 type words), these comparisons have uncovered no evidence that differences in the ability to access words as visual patterns can account for the sharp differences in spelling performance observed in this study.

Instead, the results of the spelling test suggested that linguistic factors play an important role in spelling. For both good and poor spellers the accuracy with which words were spelled was clearly influenced by the variations in orthographic transparency represented by the three levels of words. Spelling was most accurate in cases where the underlying morphophonemic structure was straightforwardly reflected in the phonetic realization of the word and became progressively more difficult as the relationship between the underlying morphophonemic structure and the written representation became increasingly obscured by intervening phonologic and orthographic rules.

Further evidence that linguistic abilities are critical in differentiating good and poor spellers came from the finding that the two groups were most readily distinguished by their performance on Level 2 words. If rote memory were the critical skill in spelling, Level 3 words should have most sharply distinguished the groups. Indeed, further analysis of the Level 2 errors revealed that poor spellers were less proficient in accessing the underlying morphophonemic structure when it was not clearly reflected in the phonetic realization of the word. This finding underscores what may be an important difference between the two groups: while good spellers found the spelling of words involving access to morphophonemic structure significantly easier than words involving the implementation of orthographic conventions, poor spellers did not.

Experiment 2

The primary purpose of Experiment 2 was to discover whether the abilities that underlie spelling competence are instances of specific learning or whether they are generalizations that can be applied productively to other English words. Specifically, the question addressed was whether college students who differ in their ability to spell familiar words would also differ in their ability to spell pseudowords that conform to the phonotactic constraints of English. The specific spelling skills under investigation included knowledge of the recurrent spelling patterns of English orthography, familiarity with the morphological principles guiding the use of prefixes and suffixes and ability to use morphophonemic information to disambiguate reduced vowels. The relevance of these skills to other aspects of written language, namely word recognition and reading comprehension, was also examined. A secondary purpose of the experiment was to explore the possibility that good and poor spellers differ in their ability to learn and subsequently to recognize nonlinguistic, nonrepresentational visual patterns.

Method

Subjects. The intent was to include the 15 best and the 15 poorest spellers from Experiment 1, but because some of the original subjects were unavailable for Experiment 2, eleven additional subjects were recruited from the original subject pool. The 15 spellers constituting the good speller group all scored more than one standard deviation above the mean on the earlier described Spelling Test of Experiment 1 (mean error score = 23.7); the 15 poor spellers scored at least one standard deviation below the mean

(mean = 44.1). The mean WRAT spelling grade equivalent was 13.9 for the good spellers and 10.5 for the poor spellers. Eight of the good spellers and 11 of the poor spellers had participated in Experiment 1.

Stimuli and procedure. The following tasks, designed to evaluate specific metalinguistic and nonlinguistic abilities relating to spelling, were administered. The 30 subjects were tested in small groups in two one-hour sessions.

1. *Knowledge of Abstract Spelling Patterns*. This task assessed the subjects' knowledge of the 174 principal spelling patterns identified by Hanna et al. (1966). The patterns included 93 consonant patterns and 81 spellings for vowels.

A list of 348 English-like spoken pseudowords was prepared and recorded on magnetic tape. It included two items for each of Hanna's 174 spelling patterns. Pseudowords that adhere to the phonotactic constraints of English were used instead of actual words in order to promote adoption of an analytic mode of processing; that is, to discourage the subjects from responding to items holistically as they might well do in the case of overlearned, familiar words.

Each dictated pseudoword was printed on a prepared sheet. In each a single spelling pattern was underlined. In half of the items the underlined portion constituted an acceptable spelling for the corresponding phoneme and in half an impossible spelling. In each case, the nonunderlined portion was spelled in a manner consistent with English orthographic practice. All 348 items appeared as orthographically acceptable letter sequences regardless of whether the underlined portion was appropriately spelled; that is, there were no letter sequences that do not occur in English. In those items where the underlined spelling was not a legitimate representation of the corresponding phoneme, the presented spellings were confined to the appropriate class of phoneme (consonant or vowel) but never included spelling patterns that could, in any English context, legitimately represent the targeted phoneme.

The tape-recoded stimuli were presented at intervals of six seconds. Subjects were asked to circle "yes" if the underlined portion of the stimulus word was judged to be an acceptable spelling of the target segment or to circle "no" if it was not. Three sample items were administered as a pretest.

2. *Principles of Prefixation* assessed knowledge of how the orthography attaches the prefix to the base word. A list of 60 items was prepared for auditory presentation consisting of three types of words: monomorphemic words (for example, *constable*); words with assimilated prefixes, such as those formed by the addition of the prefix /ad/ to base words beginning with *c*, *f*, *g*, *l*, *p*, *s* and *t* (for example, *accrue*, *affluence* and *aggravate*), or those formed by addition of /con/ to base words beginning with either *m*, *l* or *n* (for example, *committee*, *collateral* and *connubial*); and words with prefixes not involving consonant assimilation such as those formed by the addition of the prefixes *mis*, *dis*, *contra* and *un* (for example, *misshapen*, *dissimilar* and *contradiction*).

In order to forestall the possibility that a subject could mechanically partition the initial letters of the word as the basis for dividing the prefix from the stem, without examining the whole word, an effort was made to include words in the list that began with the same phonetic sequence even though different principles of prefixation are involved (e.g., *constable*, *connubial*, *concurrent*).

The tape-recorded words were presented at 10-s intervals. Subjects were asked to print each dictated word and to separate the prefix from the base by a dash. They were cautioned that some of the words would not involve a prefix, in which case they were to write a dash first, followed by the spelling of the word. Three examples, with and without prefixes, were given. Items were scored correct if the letter immediately preceding and succeeding the dash was accurate.

3. *Disambiguating Reduced Vowels*. This task tested ability to access and utilize phonological information in representing reduced vowels. The test list was made up of 50 English-like words all of which ended in the unstressed syllables, /ə/*ble* or /ə/*nts*. In some cases the target pseudoword was dictated alone, while in other cases it was preceded by one or more pseudowords phonologically related to the target. In either case, relevant phonological cues were available to assist the speller in disambiguating the reduced vowel in the targeted word. For some of the items the cue was in the relationship of the spoken pseudoword to its "derivative form." The basis of the derivations is, of course, by analogy to actual words of similar structure. For example, given the strings [ɛkstrʌpt, ɛkstrʌpʃən, ɛkstrʌptəbəl], the relationship of [ɛkstrʌptəbəl] to [ɛkstrʌpt] and [ɛkstrʌpʃən] signals the use of the vowel *i* to orthographically represent the reduced vowel in the penultimate syllable of *extruptible* as in the case of the words *corrupt*, *corruption*, *corruptible*. In other cases, the phonemic context supplied by the pseudoword itself provided the necessary cue for choosing the correct spelling pattern to represent the reduced vowel. For example, the orthographic representation for the reduced vowel in the penultimate syllable of [kəntrəmIsəbəl] is most likely to be *i* since the pseudoword was formed in analogous fashion from a stem originally occurring in Latin adjectives ending in *ibilis* and later borrowed by English.

Spellings corresponding to each of the tape-recorded target pseudowords were listed, but with omission of the reduced vowel in either the final or the penultimate syllable. The omitted vowel was marked by a blank space in the appropriate location. Beside each pseudoword, two vowel spellings were presented as choices, *a* and *i* for pseudowords ending in /ə/*ble* and *a* and *e* for items ending in /ə/*nts*. The subject's job was to choose the correct spelling for the reduced vowel.

4. *Principles of Suffixation*. To assess mastery of the principles for appending suffixes, a list of 24 pseudowords was prepared for taped presentation along with directions for changing each word into a new word by adding a given suffix. Thirteen English orthographic "rules" were incorporated (for a listing of the rules see Witherspoon, 1973, p. 282-285).

The items were dictated at 10-s intervals in a standard carrier phrase, which instructed the subjects to change each stimulus item to a related form by attaching a specified suffix (for example, "Change *prin* to *prinnish*"). The answer sheet presented a spelled out version of each pseudoword with space alongside to write the word with the appended suffix.

In addition to the foregoing tasks that were specially prepared for this study several standard tests were also administered.

5. Wechsler Adult Intelligence Scale (WAIS) Vocabulary Subtest (Wechsler, 1958).

Subjects were given answer booklets in which items were printed with a space provided for the subject to write the definition of each stimulus word. Before beginning the task, the examiner read each of the stimulus words aloud.

6. WRAT Reading Recognition. Oral reading level was assessed using the reading section of the Wide Range Achievement Test (Jastak et al., 1965). This requires subjects to read aloud a series of progressively more difficult words within a prescribed time limit. The test was administered individually to each subject according to the standard procedure.

7. Scholastic Aptitude Test Verbal Ability: (Educational Testing Service). SAT scores, required for admission to the university, were available with the subjects' permission.

8. Kimura Recurring Figures Test (Kimura, 1963). A test of memory for abstract designs that do not lend themselves readily to verbal labeling was used to assess visual memory ability. The test was chosen to provide a measure of visual memory, uncontaminated by verbal cues.

The test was administered in the standard manner. Subjects first viewed a set of 10 cards on each of which was displayed a single design. They then were shown 7 additional sets of 10 cards each. In each of the latter sets, four of the designs from the original set recur, randomly interspersed with six non-recurring designs. The task was to identify the recurring figures in each of the seven sets of cards by circling "yes" or "no" on the accompanying answer sheet.

Results and Discussion

Performance on linguistic tasks that pertain to spelling. As can be seen in Table 2 the general error pattern for the two subject groups was remarkably similar. In both groups errors on vowel patterns accounted for approximately 68 percent of the total error score while consonant errors accounted for the remaining 32 percent. But overall, the poor spellers made significantly more errors than did the good spellers in recognizing acceptable spelling patterns for English morphophonemes, $t(28) = 5.35$, $p < .001$. The greater difficulty experienced by poor spellers occurred both in identifying consonant patterns, $t(28) = 3.21$, $p < .01$, and vowel patterns, $t(28) = 5.23$, $p < .001$.

In segmenting prefixes from base morphemes, poor spellers again demonstrated significantly more difficulty than did good spellers, $t(28) = 3.81$, $p < .001$. There was no difference between good and poor spellers in segmenting nonassimilated prefixes from their base morphemes, $t(28) = 1.47$, $p > .05$, but a significant difference emerged in segmenting prefixes involving consonant assimilation, $t(28) = 3.48$, $p < .01$. The nature of the difficulty encountered by both groups was the same. Errors resulted from a failure to use the double consonant pattern at the juncture of the prefix and the base morpheme (for example, representing con-nubial as "co-nubial").

It is of interest to note that although good and poor spellers did not differ significantly in recognizing the monomorphemic words, $t(28) = 1.67$, $p > .05$, both groups found this aspect of the task difficult. Attempts to segment

---

Table 2

Summary Scores for Good and Poor Spellers on Linguistic and Nonlinguistic Tasks

	Good Spellers			Poor Spellers		
Task	Mean Error	Standard Deviation	Percent Total	Mean Error	Standard Deviation	Percent Total
1. Abstract Spelling Patterns Test						
Consonant Errors	4.7	2.9	32	8.1	2.9	32
Vowel Errors	10.0	2.5	68	17.1	4.6	68
Total Errors	14.7	4.5	--	25.1	6.1	--
2. Prefixation Test						
Nonassimilated Prefixes	2.2	1.9	15.5	3.1	1.3	14.4
Assimilated Prefixes	4.1	2.5	28.9	7.9	3.4	36.6
No Prefixes	7.9	4.8	55.6	10.6	4.2	49.1
3. Suffixation Test						
Total Errors	4.1	1.6		9.8	3.2	
4. Reduced Vowel Test						
Total Errors	9.7	3.0		18.8	3.8	
5. Kimura Figures						
Total Errors	8.8	4.7		9.3	5.2	

---

words not having prefixes (for example writing *constable* as "con-stable") accounted for approximately 50 percent of the total error score.

On the remaining linguistic tasks good spellers continued to outperform poor spellers. On the test of suffixation, poor spellers made significantly more incorrect responses than the good spellers, $t(28) = 6.08$, $p < .001$. Similarly, in representing the reduced vowel in various pseudowords, poor spellers made significantly more errors, $t(28) = 7.29$, $p < .001$.

In contrast to the sharp differences between the groups on the tasks assessing linguistic ability, no difference in the performance of good and poor spellers was found on the visual memory task, $t(28) = 0.30$, $p > .05$. This finding suggests that while the ability to remember visual information may enhance spelling proficiency in some individuals, it may not by itself account for the performance differences observed in this sample of college students.

Performance on reading and vocabulary tasks. It was also of interest to determine whether the two groups of university students could be distinguished on tests of reading ability. Whereas both good and poor spellers demonstrated college level proficiency in reading English words and in verbal scholastic aptitude, good spellers were distinctly superior to poor spellers in both these areas. As shown in Table 3 on the reading subtest of the WRAT good spellers obtained a mean grade equivalent score two years above that achieved by the poor spellers (15.3 years versus 13.3 years, respectively). Differ-

Table 3

Summary Scores for Good and Poor Spellers on Reading and Vocabulary Measures

Measure	Good Spellers Mean Score	Good Spellers Standard Deviation	Poor Spellers Mean Score	Poor Spellers Standard Deviation
1. WRAT Reading Grade Equivalent	15.3	1.3	13.3	1.7
2. Scholastic Aptitude Test Verbal Aptitude	534	75.8	465	66.7
3. WAIS Vocabulary Subtest Scaled Score	14.6	1.8	13.5	1.5

ences between the groups in reading ability were found both on the WRAT test of oral reading, $t(28) = 3.49$, $p < .002$, and on comprehension of printed text as assessed by the verbal aptitude score on the Scholastic Aptitude Test, $t(28) = 2.57$, $p < .01$. Together these results suggest that the linguistic abilities associated with differences in spelling proficiency may also contribute to differences in broader aspects of skill in written language. The fact that reading ability, as it was assessed on these two measures, was less conspicuously retarded than the spelling performance of the poor spelling group may stem from the fact that reading is a recognition task and, as such, provides more opportunities than are available in spelling for arriving at the correct answer by using contextual cues. The easier demands made by reading may therefore mask the difficulties that more readily surface in written language tasks requiring production.

In contrast, it is notable that no reliable difference between the groups was obtained on the WAIS Vocabulary Subtest, $t(28) = 1.92$, $p > .05$. This finding suggests that performance on a measure commonly used to assess verbal intelligence is not a factor associated with differences in spelling proficiency. Instead, the findings point to a deficiency on the part of poor spellers in ability to apprehend the internal structure of words.

As anticipated, the results revealed that good spellers were consistently more sensitive than poor spellers to the structural principles embodied in the English-like pseudowords. Not only were good spellers significantly better in recognizing acceptable spelling patterns for English morphophonemes, they were also more proficient in appending both prefixes and suffixes to words and in using morphophonemic information to correctly represent phonetically neutral, reduced vowels. The finding that good spellers were able to derive the correct spelling for the pseudowords suggests that their earlier success in spelling the real words on the Experimental Spelling Test was not entirely the result of whatever ability they might have to memorize the spellings of specific words. Indeed, it would seem more reasonable to suppose that good spellers have succeeded in abstracting regularities that are instanced in the orthography and have learned to exploit this knowledge when called upon to spell. This finding is consistent with the results of a few studies that have addressed this question (Fowler, Liberman, & Shankweiler, 1977; Schwartz & Doehring, 1977). The fact that poor spellers performed as poorly on the abstract spelling tasks as they did on the familiar words of the first experiment suggests that they are either less sensitive than good spellers to the uniformities that underlie English orthography or are less apt than good spellers to access this knowledge in transcribing words.

General Discussion

The misspellings of college students provide insight into the nature of spelling difficulty and offer a means for identifying those abilities that underlie competence in spelling English words. The findings of this investigation suggest that sensitivity to linguistic structure is a critical component of spelling proficiency and may account for much of the variation between otherwise literate adults who differ in spelling achievement. The data presented here revealed that college-level students who differed greatly in spelling proficiency also differed in their sensitivity to various regularities of word structure. Poor spellers were not only less able than good spellers to abstract the orthographic regularities existing at the surface phonetic level of language, but were also less successful in penetrating below the phonetic surface of words to the underlying morphophonemic representations that are captured in a word's written form. Indeed, it was the ability to access and utilize morphophonemic knowledge, both in spelling actual words and in spelling English-like pseudowords, that most clearly differentiated good and poor spellers. The finding that these performance differences are found with pseudowords implies that the knowledge that contributes to linguistic sensitivity is of a generalized sort that can be applied to new words.

It was apparent in questioning good spellers that their linguistic sensitivity was often not manifested in an explicit form that could be verbalized. Although, in some instances, individuals could describe the principles underlying their choice of a particular spelling pattern, in many other instances they were unable to explain how their choices were made. This suggests that linguistic sensitivity involves tacit knowledge as well as a more explicit understanding of how written language maps onto its spoken form. By exploiting this knowledge good spellers were able to avoid many pitfalls in spelling that proved to be insurmountable to subjects lacking in this sensitivity, as for example, the representation of reduced vowels and the affixation of prefixes and suffixes to base morphemes.

This investigation suggests that some college students have inadequately learned the principles by which writing represents the language, despite the lack of apparent deficits in reading. Of course, it is not surprising that reading would be easier than spelling, since reading is a recognition task that provides multiple cues and requires only a passive recognition of spelling patterns.

The possibility exists that some poor spellers may be experiencing difficulty not because they are insensitive to the various kinds of regularities existing at different levels of linguistic structure, but because they fail to apply this knowledge in spelling. It would be of interest to determine whether poor spellers could appreciably improve their spelling accuracy after receiving some instruction about how their linguistic competence might assist them in deriving the orthographic representation of words.

It is, of course, unlikely that differential access to linguistic structure can account for all variations in spelling proficiency. Other investigators have found spelling difficulties in some individuals to be associated with underlying deficits in serial ordering ability (Kinsbourne & Warrington, 1964; Orton, 1937; Lecours, 1966) or with dysfunctions in aspects of visual or auditory perception (Critchley, 1970; Boder, 1973). However, these investigations were conducted either on children with developmental dyslexia or on adults with acquired dyslexia following brain damage. Therefore the findings of these studies may be of limited relevance to the questions with which this study is concerned. Although some writers have proposed that individual variation in the spelling proficiency of adults is largely the result of differences in visual memory (Shaw, 1965; Witherspoon, 1973), no evidence of differences related to visual memory was found among the good and poor spellers in this study.

At all events, it is clear that competence in spelling involves more than rote memorization of words. It requires the ability to abstract regularities instanced in word structure at several levels of representation. At the most basic level, it entails abstracting the spelling patterns that stand in approximate correspondence to the phonemes of English. At the morphemic level, it requires learning English morphemes and the conventions for combining morphemes to form new words. At a higher level, it entails learning the phonological rules that map underlying morphophonemic segments to their surface phonetic form. The latter abilities especially are critical for productive use of the orthography and seem to be lacking in many otherwise literate adults who are unable to spell proficiently.

The findings of this investigation serve to emphasize that spelling is not a skill that is fully acquired as a part of an elementary education. Many young adults continuing on in higher education have persistent spelling problems. This study has produced evidence that spelling is not an isolated, low-level ability, but, like other aspects of writing skill, draws upon a variety of linguistic abilities, which continue to develop with experience, and which may be poorly developed even in highly-selected college students. The findings reported here would seem to lend substance to the claim (Chomsky, 1970) that some abilities required for full use of an alphabet are rather late intellectual developments.

References

Barron, R. W. (1980) Visual and phonological strategies in reading and spelling. In U. Frith (Ed.), Cognitive processes in spelling. London: Academic Press.

Boder, E. (1973). Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. Developmental Medicine & Child Neurology, 15, 663-687.

Chomsky, N. (1970). Phonology and reading. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 12, 335-359.

Critchley, M. (1970). The dyslexic child. Springfield, IL: Charles C. Thomas.

Educational Testing Service, Scholastic Aptitude Test.

Fischer, F. W. (1980). Spelling proficiency and sensitivity to linguistic structures. Unpublished doctoral dissertation, University of Connecticut.

Fowler, C. A., Liberman, I. Y., & Shankweiler, D. (1977). On interpreting the error pattern of the beginning reader. Language and Speech, 20, 162-173.

Fowler, C. A., Shankweiler, D., & Liberman, I. Y. (1979). Apprehending spelling patterns for vowels: A developmental study. Language and Speech, 22, 243-252.

Frith, U. (1978). Spelling difficulties. Journal of Child Psychology and Psychiatry, 19, 279-285.

Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H. Jr. (1966). Phoneme-grapheme correspondences as cues to spelling improvement. Washington, DC: U.S. Government Printing Office.

Jastak, J., Bijou, S. W., & Jastak, S. R. (1965). Wide range achievement test. Wilmington, DE: Guidance Associates.

Kimura, D. (1963). Right temporal lobe damage. Archives of Neurology, 8, 264-271.

Kinsbourne, M., & Warrington, E. (1964). Disorders of spelling. Journal of Neurology, Neurosurgery and Psychiatry, 27, 224-228.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American Engish*. Providence, RI: Brown University Press.

Lecours, A. R. (1966). Serial order in writing--A study of misspelled words in "developmental dysgraphia." *Neuropsychologia*, *4*, 221-241.

Lewis, N. (1962). *Dictionary of correct spelling*. New York: Funk & Wagnalls.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, *18*, 201-212.

Marcel, T. (1980). Phonological awareness and phonological representation: Investigation of a specific spelling problem. In U. Frith (Ed.), *Cognitive processes in spelling*. London: Academic Press.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1978). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323-331.

Orton, S. T. (1937). *Reading, writing and speech problems in children*. New York: Norton.

Perfetti, C., & McCutchen, D. (in press). Speech processes in reading. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 7). New York: Academic Press.

Schwartz, S., & Doehring, D. (1977). A developmental study of children's ability to acquire knowledge of spelling patterns. *Developmental Psychology*, *13*, 419-420.

Shaw, H. (1965). *Spell it right*. New York: Barnes and Noble, Inc.

Sloboda, J. A. (1980). Visual imagery and individual differences in spelling. In U. Frith (Ed.), *Cognitive processes in spelling*. London: Academic Press.

Steinberg, D. D. (1973). Phonology, reading, and Chomsky and Halle's optimal orthography. *Journal of Psycholinguistic Research*, 1973, *2*, 239-258.

Tenney, Y. J. (1980). Visual factors in spelling. In U. Frith (Ed.), *Cognitive processes in spelling*. London: Academic Press.

Vellutino, F. (1979). *Dyslexia: Theory and research*. Cambridge, MA: MIT Press.

Webster's seventh new collegiate dictionary. (1963). Springfield, MA: G. & C. Merriam Co.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence*. Baltimore, MD: The Williams & Wilkins Co.

Witherspoon, A. (1973). *Common errors in English*. New Jersey: Littlefield, Adams & Co.

Footnotes

[1] The nature of the mapping between phonemes and their graphemic representations is the subject of considerable debate, particularly in the case of the so-called "silent" letters. Whereas some silent letters (such as the *e* in make, life, and code) function as diacritic markers for a preceding vowel phoneme and as such may readily be classified as part of the vowel spelling, others serve no obvious function (e.g., the *b* in lam*b* or the *u* in g*u*ard). In such instances it is not clear with which phoneme the grapheme is to be associated. We have followed Hanna et al. (1966) in classifying "silent" consonant graphemes with consonant phonemes and "silent" vowel graphemes with vowel phonemes. According to this procedure the *gn* in gnaw is treated as a single spelling pattern. Thus, any of the following spellings for /n/ would be scored as substitution errors (nn, kn, pn, mn) since like gn they are alternative spelling patterns for /n/.

Appendix 1

ORTHOGRAPHIC AND MORPHOPHONEMIC GENERALIZATIONS EXEMPLIFIED IN THE LEVEL 2A AND 2B WORDS USED IN EXPERIMENT 1

LEVEL 2A

1. Words of one syllable ending in a single consonant that follows a single vowel double the final consonant before a suffix beginning with a vowel. Examples include: clannish, strapped, sobbing, and thinned (Witherspoon, 1973, p. 282).

2. Words ending in silent e usually drop the e before a suffix beginning with a vowel. However, words ending in ce and ge, and a few other words, do not drop the silent e before a suffix beginning with certain vowels. Examples include: changeable and noticeable (Witherspoon, 1973, p. 282).

3. Words ending in silent e preceded by one or more consonants usually retain the e before a suffix beginning with a consonant. Examples include: sincerely, ninety, and definitely (Witherspoon, 1973, p. 283).

4. In American usage, the final e is usually dropped before the suffix -ment when it is preceded by dg. An example is abridgment (Witherspoon, 1973, p. 284).

5. Final y following one or more consonants changes to i before the addition of letters other than i. Examples includes: flier and skies (Witherspoon, 1973, p. 284).

6. Words ending in c add k before an additional syllable beginning with e, i, or y. An example is picnickers (Witherspoon, 1973, p. 284).

7. In combinations with ful the second l of the word full is dropped when the word is used as a suffix. An example is skillful (Witherspoon, 1973, p. 285).

8. I before e except after c, or when sounded as A, as in neighbor or weigh. Examples include: disbelieve, beige, and unperceived (Witherspoon, 1973, p. 276).

9. Some nouns ending in o preceded by a consonant add es to form the plural. Others, including most musical terms that end in o, add s to form the plural. An example is echoes (Witherspoon, 1973, p. 294).

LEVEL 2B

1. Words of more than one syllable, ending in a single consonant preceded by a single vowel, if accented on the last syllable usually double the final consonant before a suffix beginning with a vowel. Examples include: preferring, omitted, equipped, and regrettable (Witherspoon, 1973, p. 282).

2. When a prefix ends with the same letter with which the root to which it is to be united begins, retain both letters in spelling the word. Examples include: misspell and dissimilar (Witherspoon, 1973, p. 277).

3. When the prefix /ad/ is appended to base words beginning with the letters c, f, g, l, p, s, or t, the d is assimilated and is orthographically represented by the letter beginning the base word. An example is aggravate (Webster, 1963, p. 10).

4. When the prefix /con/ is appended to base words beginning with either m, l, or n, the n is assimilated and is orthographically represented by the letter beginning the base word. Examples include: commemorate and commiserate (Webster, 1963, p. 164).

5. The identity of reduced vowels within words can often be recovered by relating the word to cognate forms in which the same vowel segment is not reduced. Examples include: grammar - grammatical, continuance - continuation, inspiration - inspire, repetition - repeat.

6. If the root forms its noun by the immediate addition of -ion, the correct ending is likely to be ible. There are, however, exceptions. Examples include: indigestible and inexhaustible (Lewis, 1962, p. 103).

7. If the root ends in -ns, the ending is probably -ible. An example is defensible (Lewis, 1962, p. 103).

8. If the root to which the suffix is to be added is a full word in its own right, the correct ending is usually able. An example is regrettable (Lewis, 1962, p. 1).

9. If a two-syllable verb ending in -er is accented on the first syllable, the noun ending is likely to be -ance. An example is utterance (Lewis, 1962, p. 13).

10. If a verb ends in -ear, the likely ending is ance. An example is clearance (Lewis, 1962, p. 13).

Appendix 2

WORD LIST

LEVEL 1	LEVEL 2A	LEVEL 3
1. yam	1. strapped	1. chihuahua
2. inflate	2. skillful	2. onomatopoeia
3. adverb	3. cancelled	3. Fahrenheit
4. vortex	4. picnickers	4. plagiarism
5. cameo	5. abridgment	5. sarsaparilla
6. harp	6. flier	6. hemorrhage
7. terminates	7. changeable	7. sergeant
8. trump	8. sincerely	8. eunuch
9. vacate	9. echoes	9. connoisseur
10. update	10. disbelieve	10. mnemonic
11. vibrated	11. sobbing	11. reveille
12. mandated	12. beige	12. desiccate

13. compensates
14. delimit
15. zebra
16. blunder
17. emit
18. boxer
19. repent
20. intertwined

13. skies
14. unperceived
15. clannish
16. noticeable
17. ninety
18. thinned
19. basically
20. definitely

13. syphilis
14. pygmy
15. sacrilegious
16. diphtheria
17. hieroglyphic
18. thumb
19. gnaw
20. lengthen

LEVEL 2B

21. uncover
22. diplomat
23. retort
24. canister
25. clustering
26. undiminished
27. terminology
28. mask
29. manifestation
30. definitions
31. frustrated
32. expectation
33. alternate
34. stimulation
35. examiner
36. preventive
37. unemployment
38. punishment
39. establishing
40. electronics

1. misspell
2. aggravate
3. commemorate
4. defensible
5. grammar
6. clearance
7. inexhaustible
8. utterance
9. continuance
10. prevalent
11. dissimilar
12. preferring
13. inspiration
14. omitted
15. repetition
16. indigestible
17. recommend
18. regrettable
19. equipped
20. commiserate

21. Wednesday
22. soldered
23. talker
24. subpoena
25. annihilate
26. rhododendron
27. kaleidoscope
28. pyorrhea
29. bourgeois
30. thigh
31. listener
32. slaughter
33. indebted
34. climb
35. answering
36. knock
37. beautifully
38. laugh
39. folk
40. tongue

EFFECTS OF PHONOLOGICAL AMBIGUITY ON BEGINNING READERS OF SERBO-CROATIAN*

Laurie B. Feldman,† G. Lukatela,†† and M. T. Turvey†††

Abstract. Third- and fifth-grade Yugoslavian children were tested on rapid naming of familiar words and unfamiliar pseudowords that were a) written in either the Roman alphabet or the Cyrillic alphabet and b) were either phonologically ambiguous or not. Phonological ambiguity was produced by using letter strings that, when transcribed in Roman or when transcribed in Cyrillic, contained one or more ambiguous characters. Ambiguous characters are those letters shared by the two alphabets that receive different phonemic interpretations in the two alphabets. The controls for phonologically ambiguous words were the same words in their alternative, non-ambiguous alphabetic transcription. Consistent with previous experiments on adults, the phonologically ambiguous form of a word or pseudoword was named much more slowly than the phonologically unambiguous form. For children who were equally proficient in both Roman and Cyrillic, the effect of phonological ambiguity was greater as children named letter strings faster. If it can be assumed that reading fluency correlates with naming latency, then it can be argued that the better beginning reader is more phonologically analytic.

The present paper reports an experiment on the rapid naming of printed letter strings by Yugoslavian children. In Yugoslavia, children are taught two alphabets: a Roman alphabet (the characters of which would be fairly familiar to the reader of English) and a Cyrillic alphabet (the characters of which are similar to but not identical with Russian script). Ordinarily, Yugoslavian children learn both alphabets by the end of the second grade and are reasonably proficient in both by the fifth grade. (In Belgrade, where most of the children in the present experiment were educated, the Cyrillic alphabet is taught first.) Unlike the English writing system, the two writing systems of the Serbo-Croatian language maintain strict grapheme-phoneme correspondences; the phonemic interpretation of a letter does not vary with context and there are no letters made silent by context. Nevertheless, confusion

---

*In press, Journal of Experimental Child Psychology.

†Also University of Delaware.

††University of Belgrade.

†††Also University of Connecticut.

Acknowledgment. We wish to thank Vesna Ognjenović and the students, teachers, and director of the Svetozar Miletić School in Zemun, Yugoslavia for making this work possible. In addition, we thank Milena Cicmilović, Nevena Vucić, Anne Ullrich, and Petar Makara for helping to collect and analyze the data. This work was supported by NICHD Grant HD-08495 to the University of Belgrade and by NICHD Grant HD-01994 to Haskins Laboratories.

TABLE 1

SERBO-CROATIAN				
ROMAN		CYRILLIC		
PRINTED		PRINTED		LETTER NAME IN I.P.A.
UPPER CASE	LOWER CASE	UPPER CASE	LOWER CASE	
A	a	А	а	a
B	b	Б	б	bə
C	c	Ц	ц	tsə
Č	č	Ч	ч	tʃə
Ć	ć	Ћ	ћ	tʃjə
D	d	Д	д	də
Đ	đ	Ђ	ђ	dʒjə
DŽ	dž	Џ	џ	dʒə
E	e	Е	е	ɛ
F	f	Ф	ф	fə
G	g	Г	г	gə
H	h	Х	х	xə
I	i	И	и	i
J	j	Ј	ј	jə
K	k	К	к	kə
L	l	Л	л	lə
LJ	lj	Љ	љ	ljə
M	m	М	м	mə
N	n	Н	н	nə
NJ	nj	Њ	њ	njə
O	o	О	о	ɔ
P	p	П	п	pə
R	r	Р	р	rə
S	s	С	с	sə
Š	š	Ш	ш	ʃə
T	t	Т	т	tə
U	u	У	у	u
V	v	В	в	və
Z	z	З	з	zə
Ž	ž	Ж	ж	ʒə

similar to that experienced by the beginning reader of English is experienced by the beginning reader of Serbo-Croatian (Mann, Liberman, & Shankweiler, 1980).

As noted above, the Serbo-Croatian language is written in two different alphabets, Roman and Cyrillic. The two alphabets transcribe one language and their graphemes map simply and directly onto the same set of phonemes. These two sets of graphemes are, with certain exceptions, mutually exclusive (see Table 1). Most of the Roman and Cyrillic letters are unique to their respective alphabets. However, the two alphabets share a number of letters. The phonemic interpretation of some of these shared letters is the same whether they are read as Cyrillic or as Roman graphemes; these are referred to as common letters. The remaining shared letters have two phonemic interpretations, one in the Roman reading and one in the Cyrillic reading; these are referred to as ambiguous letters (see Figure 1). Whatever their category, the individual letters of the two alphabets have phonemic interpretations (classically defined) that are virtually invariant over letter contexts. This reflects the phonologically shallow nature of the Serbo-Croatian orthography.

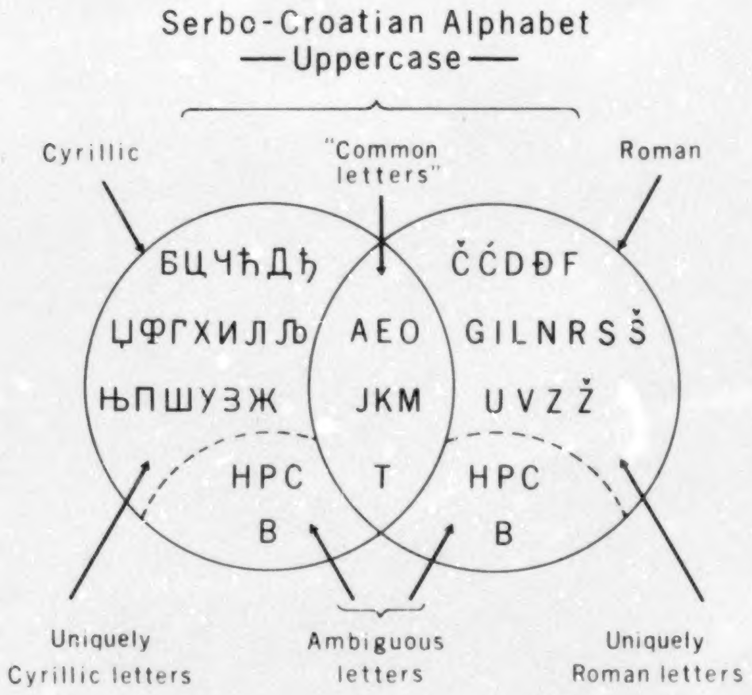


Figure 1. Letters of the Roman and Cyrillic alphabets.

The present experiment exploits this limited but explicit ambiguity in the Serbo-Croatian writing system. It does so to address the question of whether or not skilled beginning readers who have learned both the Roman and

Cyrillic alphabets can be distinguished from less skilled readers by their sensitivity to phonological ambiguity: In rapidly naming letter strings, is the better beginning reader more hampered by the presence of phonologically ambiguous characters than the poorer beginning reader? The question takes this latter form for two reasons. First, accessing the name of a letter string may entail a phonologically analytic strategy, especially when the orthography is as regular as the Serbo-Croatian orthography (Turvey, Feldman, & Lukatela, 1984). Second, facility with a phonologically analytic strategy for naming (and, more generally, for accessing the internal lexicon) may be one way to distinguish the more skilled reader from the less skilled reader.[1] Consequently, for these two reasons, it may be supposed that in Serbo-Croatian the more skilled the beginning reader the greater is his or her sensitivity to phonological ambiguity.

A similar strategy has been pursued by I. Y. Liberman, Shankweiler, and their colleagues to distinguish good and poor readers of English by their ability to use phonetic coding in the short-term retention of linguistic materials presented visually or auditorily. The general result obtained by these investigators is that good readers perform proportionately worse than poor readers when the to-be-remembered stimuli are phonetically similar compared to when they are phonetically dissimilar (Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). That is, although good readers tend to do better in short-term memory tests than poor readers, the scores of good readers are influenced more by phonetic similarity.[2]

Outside of the short-term memory task, however, evidence for a difference between good and poor readers of English that is based on a difference in sensitivity to the linguistic underpinnings of the orthography is both sparse and equivocal. For example, Barron (1978) showed that visually presented pseudohomophones (e.g., BRANE, WERD) lengthened the lexical decision latencies of good readers but not of poor readers. It is difficult, however, to draw conclusions about linguistic contributions to visual word processing on the basis of pseudohomophone effects for the following reasons. First, there is the possibility that the phonetic interpretations assigned to pseudohomophones (e.g., BRANE) and to their related words (e.g., BRAIN) may be sensitive to the orthographic differences between them. Second, even if a pseudohomophone and its related word were assigned identical phonetic interpretations, it does not mean that they would be assigned identical phonological interpretations. (In formal linguistics, the phonetic and phonological representations of an English word are distinct.) Third, it is frequently the case that the pseudohomophones used in experiments are visually less similar to English words (i.e., orthographically less well structured) than are the control pseudowords (Martin, 1982).

Speaking more generally, reliable demonstrations of a linguistic contribution (e.g., phonological) to visual lexical access with English materials have proven hard to come by, regardless of the age and fluency of the reader. This fact has been interpreted to mean that accessing the lexical representation or name of printed English words is ordinarily a linguistically nonanalytic process, often termed visual (e.g., Coltheart, 1978). Alternatively, it could be interpreted to mean that, within the confines of the experimental procedure for studying lexical access and naming, it is difficult to find a manipulation of English stimulus materials that consistently reveals a linguistic contribution.

Results of research on lexical access and naming with the Serbo-Croatian language contrast sharply with the results of research with English. It has been shown repeatedly that in tasks where the lexical status of a letter string has to be provided rapidly, the presence of ambiguous letters has a retarding effect. A phonological contribution is consistently implicated (Feldman, 1983). The basic experimental procedure has been to compare two kinds of letter strings: (1) phonologically unambiguous letter strings, comprised of letters unique to an alphabet as well as letters shared by the two alphabets (see Figure 1). (2) phonologically ambiguous letter strings, comprised solely of letters shared by the two alphabets and always including one or more ambiguous letters. The first kind of letter string can be read in only one way and has a single morphophonological representation. In contrast, the second kind of letter string can be read in two ways because it is written in the letters shared by the two alphabets, some of which are phonemically bivalent; a letter string of this kind has two distinct morphophonological representations.[3] If lexical access and naming proceed with reference to the phonology, then a phonologically ambiguous letter string might be expected to extend response time relative to a letter string that receives a unique morphophonological representation. This hypothesis has been evaluated in two ways: via a comparison of different letter strings (Lukatela, Popadić, Ognjenović, & Turvey, 1980; Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) and via a comparison of different versions (Roman and Cyrillic) of the same letter string (Feldman, 1981; Feldman, Kostić, Lukatela, & Turvey, 1983; Feldman & Turvey, 1983).

When different words are compared, problems of matching the words on frequency of occurrence in the language, richness of meaning, length, number of syllables, etc. arise. These problems can be virtually eliminated by taking advantage of the fact that some Serbo-Croatian words can be transcribed in the Roman and Cyrillic alphabets such that in one alphabet the reading is phonologically ambiguous, whereas in the other alphabet, the reading is phonologically unique. To evaluate the phonological contribution to lexical access and naming, the bialphabetical nature of Serbo-Croatian permits a comparison of a written word with itself.

Consider the Serbo-Croatian word for savanna. This word is phonologically bivalent when transcribed in Cyrillic (CABAHA, where C, B, and H are ambiguous) and phonologically unique when transcribed in Roman (SAVANA). The expectations that lexical decisions on, and the naming of, letter strings like CABAHA should be significantly slower than the same responses to letter strings like SAVANA has been confirmed experimentally (Feldman, 1981; Feldman et al., 1983; Feldman & Turvey, 1983). To reiterate, the letter strings CABAHA and SAVANA are the same word and, therefore, identical in all respects but one, namely, the number of morphophonological representations. It is, therefore, a noteworthy empirical observation that their associated latencies should differ by hundreds of milliseconds.

The design used in the present experiment with children was modeled after that used in the experiments with adults by Feldman and her colleagues (see above). Because mastery of both the Roman and Cyrillic alphabets is an essential prerequisite for the appreciation of bivalence, children were tested at two levels of alphabetic proficiency: 6 months and 30 months after they had learned the second alphabet. All children were tested on words and pseudowords that were phonologically ambiguous when transcribed in one of the two alphabets. The children's naming latencies and erroneous responses to these ambiguously transcribed letter strings and to their unambiguously transcribed controls were compared. In the experiment, the question posed above about phonological ambiguity and beginning readers of Serbo-Croatian took the form: With alphabetic proficiency controlled, is the latency (and/or error) difference between naming ambiguous and unambiguous versions of the same word (that is, the effect of phonological ambiguity) larger for the child whose reading skills are superior?

Method

Subjects

In order to include a range of reading ability at two levels of alphabetic proficiency, third- and fifth-grade students from the Svetozar Miletić School in Zemun, a suburb of Belgrade, participated in the study. The sample consisted of two complete classes at each grade level. As is the practice in Yugoslavia, these classes were not grouped by reading ability. Based on their own accounts, 85% of the children had learned the Cyrillic alphabet in the first grade and the Roman alphabet in the second. For the remaining 15%, the order of acquisition was reversed. When asked to write out their name, 95% of third graders and 73% of fifth graders chose to write in Cyrillic. Initially, 40 third graders and 37 fifth graders were tested. Three students were eliminated from the study because they often hesitated and triggered the voice key before actually initiating articulation. Two students were eliminated due to a preponderance of technical errors. Another four students were randomly eliminated in order to yield an equal number of subjects in each condition. Data from 34 students at each grade level were included in the analysis.

Materials

Two sets of letter strings were presented to each child. These included a pretest composed of 20 orthographically regular and unambiguous pseudowords all written in Cyrillic. After a brief pause, this was followed by a mixed test list. The test included 40 words and 40 pseudowords. Half of the letter strings were ambiguous and half were unambiguous. Among the ambiguous words, half were words by their Roman reading (and pseudowords by their Cyrillic reading), e.g., BA^{T}AK, and half were words by their Cyrillic reading (and pseudowords by their Roman reading), e.g., EKCEP. Among the ambiguous pseudowords, both alphabet readings were phonologically acceptable but meaningless.

Stimulus items were constructed so that each word and pseudoword could be written in two forms: A phonologically ambiguous form and the unique alphabet transcription of that same word. For the ambiguous words, half of the unique alphabet transcriptions were in Roman and half in Cyrillic. Analogously for the ambiguous pseudowords, half of the unique alphabet transcriptions were in Roman and half in Cyrillic. For the ambiguous pseudowords, however, the unique alphabet transcription was arbitrarily designated because there was no preferred phonological interpretation based on lexicality to which it need correspond.[4] In summary, there were four types of words (ambiguous/pure x Roman/Cyrillic) and three types of pseudowords (ambiguous/pure Roman/pure Cyrillic). Each child viewed words and pseudowords of each type and different forms of the same item were presented to different groups of children (see Table 2). Examples of the 40 words and 40 pseudowords and their distribution across groups is summarized in Table 2.

Table 2

Examples of AMBIGUOUS and UNIQUE letter strings
and their distribution across groups of subjects.

Lexicality	Alphabet	Phonology	Form Group 1		Form Group 2	
WORD	ROMAN	Ambiguous	BATAK	drumstick	KOBAC	hawk
		Unique	EKSER	nail	VETAR	wind
	CYRILLIC	Ambiguous	BETAP	wind	EKCEP	nail
		Unique	КОБАЦ	hawk	БАТАК	drumstick
PSEUDOWORD	ROMAN	Ambiguous	BOPAM*		HABOT	
		Unique	ROJOS		SEMON	
	CYRILLIC	Ambiguous	CEMOH*		POJOC	
		Unique	ХАБОТ		БОПАМ	

*Classification of these letter strings distinguish only in the randomly assigned alphabet of their unique alphabet transcription (see text).

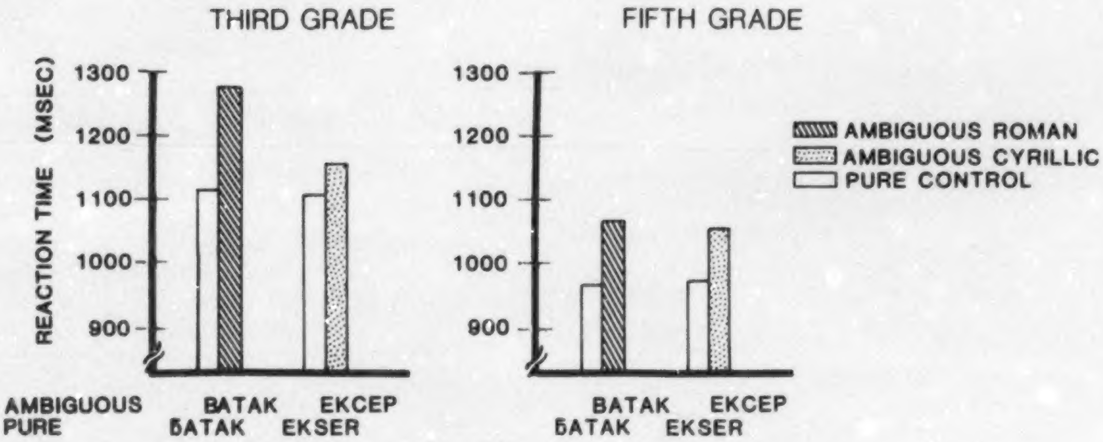


Figure 2. Mean reaction time for third and fifth graders to name AMBIGUOUS (Roman and Cyrillic) words and the UNAMBIGUOUS alphabet transcription of the same words.

All letter strings had between three and five letters and the proportion of items with three, four, and five letters was balanced across words and pseudowords in the test list. All letter strings in the pseudoword pretest contained four or five letters. All words in the test list were familiar to third- as well as fifth-grade students as judged by their teachers' assessment and by a frequency count based on children's texts in Serbo-Croatian (Lukić, 1970).

Procedure

The children performed a naming task on two lists of items, a pretest and a test. They read each word aloud as it appeared, projected onto a screen placed about 1 m in front of the child. Reaction time was measured from stimulus onset by a voice key. One experimenter recorded latencies and marked errors while a second experimenter noted the errors in more detail. In the pretest, children were instructed to read each letter string as accurately as possible. They were told that all items were pseudowords composed of four or five letters, printed in Cyrillic. After the pretest, instructions were modified to stress speed as well as accuracy. The children were also informed that the next list would be composed both of meaningful words and of letter strings that had no meaning. Further, they were clued that some of these would be printed in Cyrillic and some in Roman. Finally, they were instructed at the outset and prompted through the course of the test list to read ambiguous words by their word reading when one existed, i.e., to read BATAK as /batak/ meaning "drumstick," not as /vatak/, which is meaningless. (Only the word readings were treated as correct responses.) In summary, the ambiguous and unique forms of each word were distributed across two groups of subjects so that no subject saw two forms of the same word but all subjects saw ten ambiguous forms and ten unique forms in each alphabet. Pseudowords were designed in an analogous manner although there was no real distinction between the two alphabets for ambiguous pseudowords. Finally, practice items occurred at the beginning of both the pretest and the test list.

Results

All correct reaction times were included in the analysis of variance. Because there was a high proportion of slow latencies, some as long as 4000 ms, median reaction times were entered into the analysis of variance. Separate analyses were performed on the word and pseudoword data. In order to capture any pattern revealed by the extended latencies, a second set of error analyses was performed including incorrect responses and correct responses that were slower than 2500 ms. For both the median and the error data on words, analyses based on subject variability and on item variability (in parentheses) are reported.

The analysis of median reaction times for words revealed significant main effects for three variables: Grade, Alphabet, and Phonology. Inspection of the means in Figure 2 shows a significant effect of grade--that is third graders were slower than fifth graders, $F(1,66) = 5.58$, MSe = 281883.0, $p < .05$ ($F(1,38) = 44.60$, MSe = 15329.8, $p < .001$). In addition, there was a significant effect of alphabet--Cyrillic words were named faster than Roman words, $F(1,66) = 6.04$, MSe = 13753, $p < .05$ ($F(1,38) = 0.74$, MSe = 492078, $p < .60$). Most important, there was a significant effect of phonology--ambiguous words were slower than the unique alphabet transcription of the same word, $F(1,66) = 38.56$, MSe = 16175.9, $p < .001$, ($F(1,38) = 13.03$, MSe = 35081.2, $p <$

.001). In the subjects analysis (but not in the items analysis) 2 two-way interactions approached significance: The interaction of phonology by alphabet suggested that overall, the effect of ambiguous phonology might have been more robust in Roman than in Cyrillic, $F(1,66) = 3.37$, MSe = 20080.2, $p < .06$. And, means for the interaction of grade by alphabet indicated that third graders were slower on all Roman print than on all Cyrillic but that fifth graders read aloud comparably in both, $F(1,66) = 3.84$, MSe = 13753.8, $p < .05$. The three-way interaction of alphabet by phonology by grade missed significance, however.

The analysis of variance on incorrect and slow responses to words provided a pattern similar to that for reaction time. Third graders performed less well than fifth graders, $F(1,66) = 7.90$, Mse = 2.0879, $p < .01$ ($F(1,38) = 6.40$, MSe = .9993, $p < .05$). Ambiguous words were more likely to elicit incorrect responses than unambiguous words $F(1,66) = 234.75$, MSe = .9247, $p < .001$ ($F(1,38) = 52.05$, MSe = 4.9967, $p < .001$). There was, however, no main effect of alphabet. In the analysis of errors, the difference between third and fifth graders was larger for ambiguous words than for pure words as the interaction of phonology by grade indicated, $F(1,66) = 12.03$, MSe = .9247, $p < .001$ ($F(1,38) = 4.22$, MSe = 1.1598, $p < .05$). Moreover, as indicated by the interaction of alphabet by grade, third graders had a tendency to perform less well in Roman than in Cyrillic, while fifth graders showed the opposite pattern, $F(1,66) = 4.47$, MSe = 1.2523, $p < .05$ ($F(1,38) = 5.63$, MSe = .9993, $p < .05$). Finally, the three-way interaction of phonology by grade by alphabet was nearly significant, $F(1,66) = 3.47$, MSe = 1.6133, $p < .06$ ($F(1,38) = 7.78$, Mse = 1.1598, $p < .01$). The mean number of errors for each condition and grade are reported in Table 3.

Table 3

Mean number of errors (and standard deviation) of response latencies for ambiguous and unique forms of words in each alphabet.

	CYRILLIC AMBIGUOUS (EKCEP)	UNIQUE CONTROL (EKSER)	DIFFERENCE	ROMAN AMBIGUOUS (BATAK)	UNIQUE CONTROL (БATAK)	DIFFERENCE
THIRD GRADE	2.32[a] (361)[b]	0.47 (293)	1.85	2.94 (410)	0.41 (368)	2.53
FIFTH GRADE	2.00 (213)	0.38 (151)	1.62	1.47 (201)	0.32 (183)	1.15

[a] Errors

[b] Standard Deviation of Latencies

The analysis of median pseudoword latencies including three types of pseudowords (ambiguous, unique Cyrillic, unique Roman) indicated a significant main effect of phonology, $\underline{F}(2,132) = 27.06$, MSe = 44990.8, $\underline{p} < .001$, and a marginally significant effect of grade whereby third graders were slower than fifth graders, $\underline{F}(1,66) = 3.59$, MSe = 465152, $\underline{p} < .06$. Mean pseudoword naming times in ms for third graders were 1308 for Cyrillic, 1286 for Roman and 1574 for ambiguous letter strings; corresponding times for fifth graders were 1204, 1097, and 1324, respectively. Post hoc tests indicated that unique Roman and unique Cyrillic forms did not differ for third graders, $\underline{F}(1,66) = .12$, but that unique Roman forms were significantly faster than unique Cyrillic forms for fifth graders, $\underline{F}(1,66) = 7.14$, $\underline{p} < .009$.

Several analyses of variance suggested that alphabetic proficiency as indexed by interactions of alphabet by grade figured prominently in the pattern of results. In general, performance of fifth graders was equivalent with Roman and Cyrillic print, while third graders displayed weaker performance with Roman than with Cyrillic. In the analyses that follow, proficiency with each of the two alphabets was not confounded with measures of reading skill; the relation between reading skill and sensitivity to phonological ambiguity (which depends on the ability to derive two phonological interpretations for a letter string--one in each alphabet) was addressed separately at two levels of bialphabetic proficiency.

The Relation of Ambiguity to Decoding Speed and Error

For each subject, the difference in naming time for ambiguous and unique words was computed separately for Roman words, for Cyrillic words, and for their combined effect. These provided indices of the effect of phonological ambiguity. In addition, the median latency on the pretest with unambiguous Cyrillic pseudowords was computed for each child. Given that naming time for individual letters and pseudowords has been shown to correlate with reading skill (Jackson, 1980; Jackson & McClelland, 1979; Perfetti & Hogaboam, 1975), the median pretest latency can serve as an index of reading proficiency. The ambiguity scores were then correlated with the pretest latencies for 33 third graders and 34 fifth graders. (A reading skill measure was missing for one third grader.) Correlations were computed separately for both grades, since grade provided an index of bialphabetic proficiency. Moreover, separate correlations insured against a correlation produced by sampling from two extreme groups because third graders were generally slower than fifth graders. The correlation between the degree to which naming was slowed down by ambiguity and reading skill (as indexed by the pseudoword naming task) was significant for Cyrillic words alone for third graders, $\underline{r} = -.430$, $\underline{p} < .05$, and nearly significant for fifth graders, $\underline{r} = -.297$, $\underline{p} < .10$. The correlation between ambiguity and reading skill for Roman words alone was nonexistent for third graders and nearly significant for fifth graders, $\underline{r} = -.274$, $\underline{p} < .20$. Finally, the correlation between ambiguity averaged over alphabets and reading skill was not significant for third graders but was significant for fifth graders, $r = -.378$, $\underline{p} < .05$. These results are summarized in Table 4. The negative correlations indicate that, in general, the faster reader is more impaired by phonological ambiguity. Overall, the effect is strongest in the Cyrillic alphabet, that is, the alphabet learned first.

Classification of errors showed that overall, third and fifth graders did not distinguish on the types of errors they made, although third graders tended to make more errors generally. Three types of errors were identified: 1)

Table 4

Correlation of reading skill with detriment due to ambiguity.

AMBIGUOUS TRANSCRIPTION	THIRD GRADE	FIFTH GRADE	COMBINED
Cyrillic	-.430**	-.297++	-.398***
Roman	-.04	-.274+	-.011
Combined	-.28+	-.378*	-.286*
	N = 33	N = 34	N = 67

+*p* < .20
++*p* < .10

**p* < .05
***p* < .02
****p* < .01

Table 5

Proportion of wrong alphabet, mixed alphabet and substitution/hesitation errors for all words by third and fifth graders.
(Numbers represent percents.)

	ERROR		
GRADE	WRONG ALPHABET	MIXED ALPHABET	SUBSTITUTION/ HESITATIONS
THIRD	12.95	3.75	3.28
FIFTH	10.8	3.05	2.18

Reading an ambiguous word in the wrong alphabet, for instance, giving BATAK a meaningless Cyrillic reading when it means "drumstick" read as Roman. Note that these errors can occur only with words. 2) Mixing alphabets within a word, e.g., reading one ambiguous character in Roman and the following character in Cyrillic. 3) Hesitating, or reversing, or substituting a different phoneme for the one that is specified. In classifying errors, a given word for a given subject was never entered into two categories. Where an error was classifiable in more than one way, wrong-alphabet and mixed-alphabet designations took priority over substitutions and hesitations, but unique word errors were necessarily of the latter variety. The error data reported below are restricted to words, both ambiguous and unique, and are summarized in Table 5.

Pure words were excluded from subsequent analyses. Separate analyses of variance were performed for wrong alphabet and mixed alphabet errors on ambiguous words. Inspection of mixed alphabet means in Table 6 and the results of the analysis indicate no significant main effects or interactions. In the wrong alphabet error analysis, there was a significant interaction of alphabet by grade, $\underline{F}(1,66) = 4.94$, $MSe = 1.520$, $\underline{p} < .05$, ($\underline{F}(1,38) = 3.09$, $MSe = 1.620$, $\underline{p} < .05$). Consistent with the latency data on alphabetic proficiency described above, third graders found the Roman ambiguous words more difficult than the Cyrillic ambiguous words, while fifth graders found them equivalent.

Table 6
Mean number of wrong-alphabet and mixed-alphabet errors
(and standard deviation) for ambiguous Roman and Cyrillic words

	THIRD GRADE		FIFTH GRADE	
AMBIGUOUS TRANSCRIPTION	WRONG ALPHABET	MIXED ALPHABET	WRONG ALPHABET	MIXED ALPHABET
CYRILLIC	.91 (1.23)	.44 (.61)	1.22 (.91)	.44 (.76)
ROMAN	1.63 (1.36)	.34 (.60)	.97 (1.09)	.19 (.29)

In order to ascertain whether type of error varied with reading skill, each type of error was correlated with scores on the pretest. From the pretest, two indices of reading skill were developed: Median pretest naming time and the number of pretest errors--substitution, hesitation, or reversals. As above, correlations were computed separately for grades. In this case, number of mixed-alphabet errors correlated significantly with reading skill. Results are summarized in Table 7. The positive correlations indicate that for readers who are equally proficient in both alphabets, less skilled decoders were more likely to mix alphabets within a word (Roman or Cyrillic) than were more skilled decoders.

240

Table 7A

Median of Decoding Latencies (standard deviations) and their correlation with Mixed-Alphabet Errors.

MEASURE	THIRD GRADE	FIFTH GRADE	THIRD/FIFTH
DECODING LATENCY (S.D.)	969(270)	878(154)	924(223)
CORRELATION	.182	.365*	.250*

Table 7B

Median Decoding Errors (standard deviation) and their correlation with Mixed-Alphabet Errors.

DECODING ERRORS (S.D.)	2.65(1.99)	2.16(2.19)	2.40(2.09)
CORRELATION	.466***	.355*	.413***
	N = 33	N = 34	N = 67

$^{*}p < .05$
$^{***}p < .01$

Discussion

If a reader named words strictly on the basis of their familiar figural aspects, then naming a familiar printed word transcribed with one or more ambiguous characters (e.g., BATAK) should not be any different from naming the same familiar word transcribed with no ambiguous characters (e.g., БATAK). It is evident, however, that phonologically ambiguous letter strings were named, in general, more slowly than were their phonologically unambiguous controls. Evidently, the readers in the present experiment did not treat words as holistic figural patterns. On the contrary, the data suggest that the beginning readers in the experiment noticed (more or less) the phonological aspects of a printed word that were specified in the details of its orthographic structure. In this latter respect the present data replicate with children the observations made previously with adults.

When bialphabetic adult readers of Serbo-Croatian performed a lexical decision task, letter strings composed of ambiguous and common characters incurred longer latencies than the unique alphabet transcription of the same word (Feldman & Turvey, 1983) and, in an analogous naming task, the same pattern of results occurred (Feldman, 1981). In the adult experiments, words were selected so as to include a varied distribution in the number and position of the ambiguous characters within the letter string. Results indicated that all letter strings that could be assigned both a Roman and a Cyrillic reading incurred longer latencies than the unique alphabet transcription of the same word and that the magnitude of the difference between the ambiguous form of a word and its unique alphabet control depended on the number and distribution of ambiguous characters in the ambiguous letter string. These results with phonologically bivalent letter strings were interpreted as evidence that both lexical decision and naming in Serbo-Croatian necessarily involve an analysis that is sensitive to phonology and component orthographic structure. Moreover, in an earlier study, words and pseudowords composed entirely of common letters (with no ambiguous or unique letters) were accepted and rejected, respectively, no more slowly than letter strings that included common and unique letters. Because the distinction between common letters and ambiguous letters was based on their phonemic interpretation, this result suggested that it was *phonological* bivalence rather than a figure-based alphabetic bivalence that governed the effect (see Lukatela et al., 1978, 1980, for a complete discussion). In summary, the adult studies suggested that processes of word recognition were both *analytic* and *phonological* in nature.

The major question of interest in the present study was whether the magnitude of the difference between the ambiguous forms and their unique alphabet controls was larger for the more fluent beginning readers: Were the better beginning readers caused to respond proportionately more slowly (and/or to commit proportionately more errors) by phonological ambiguity than the poorer beginning readers? In the present experiment, the measure of reading skill was the speed with which a reader named unfamiliar, nonsense letter strings. This speed should be faster, on the average, the better are the reader's decoding skills. However, the answer to the ambiguity question was not independent of another more general question, namely, the subjects' relative proficiency in Roman and Cyrillic. Sensitivity to phonological ambiguity necessarily entails the ability to (automatically) assign phonological interpretations in *two* alphabets (see Turvey et al., 1984)--an ability that requires approximately equal familiarity with both alphabets.

The analysis of variance on median reaction times indicated that third graders named letter strings more slowly and less accurately than fifth graders. The analysis also revealed that the performance of all children--third and fifth graders--deteriorated on ambiguous letter strings relative to their unique alphabet controls. The major question of interest, however, was whether the more skilled beginning reader was more impaired by phonological ambiguity than the less skilled beginning reader, independent of general proficiency with each alphabet. Although the interactions of phonology by grade or of phonology by grade by alphabet were not significant, the relation between phonological ambiguity and reading skill was evaluated separately for each level of alphabetic proficiency for several reasons. First, inspection of the means in Figure 2 indicates that for fifth graders the effect of phonological ambiguity was constant over alphabet, whereas for third graders the effect was more exaggerated for those letter strings that were ambiguous in the Roman alphabet. (A comparison of the variances across groups of words in Table 3 suggests the same thing.) Second, although neither of the two-way interactions was significant by the items analysis, one (alphabet by phonology) was almost significant ($\underline{p} < .06$), and the other (alphabet by grade) was significant ($\underline{p} < .05$) by the subjects analysis. Generally, the chances of obtaining higher order interactions with the dichotomous grade variable were further reduced by the magnitude of the variability in the latency data of the third-grade children. Third, the error data of the third-grade children suggested that their facility with Roman letter strings in general was not as good as their facility with Cyrillic letter strings in general, whereas no such bias was evident in the fifth-grade data.

Apparently, third graders were less proficient in the newly acquired Roman alphabet and they found it difficult to suppress the unwanted Cyrillic reading of an ambiguous Roman letter string. For third graders the first-learned and more familiar (Cyrillic) alphabet tended to dominate their naming responses. Because a Cyrillic bias would exaggerate the effect of ambiguous characters in Roman words and reduce the effect of ambiguous characters in Cyrillic words, it counters any true phonological ambiguity effect. It appears that the dominance of the first-learned alphabet has waned considerably by the fifth grade, however. The analysis on wrong alphabet errors (Table 6) supports this interpretation. It should be remarked, however, that there is some evidence to suggest an asymmetry between the first- and second-learned alphabets--with a continued dominance of the first-learned--that persists, in more difficult tasks, through adulthood (Lukatela, Savić, Ognjenović, & Turvey, 1978). The literature on interference patterns between languages using the Stroop and dichotic listening tasks (Magiste, 1984) provides evidence for a similar asymmetry. The detrimental influence of the second-learned language on the first-learned language and the influence of the first-learned language on the second-learned language is symmetric when proficiency in both languages is balanced, but asymmetric when one language is dominant. As proposed elsewhere, in terms of the interaction of two symbol systems bialphabetism may be a limited case of bilingualism (Feldman, 1983; Lukatela, Savić, Ognjenović, & Turvey, 1978).

As discussed above, speed of naming nonsense letter strings was taken as the measure of a child's reading skill. Speed of naming nonsense items was then correlated with the ambiguous-unambiguous latency difference. Larger differences were associated with faster naming times (Table 4). That is, faster decoders were slowed proportionately more by phonological ambiguity than slower decoders. Examination of the sub-correlations revealed that this significant correlation was carried in largest part by the difference between words that were phonologically ambiguous in Cyrillic and their unique alphabet controls. The difference between words transcribed ambiguously in Roman and their unambiguous controls did not correlate significantly with the decoding speeds of third graders (see Table 4). However, as noted above, there was considerable variance in individual ambiguity scores for Roman materials, suggesting an inconsistency in the ability of some third graders to handle letter strings in the newly acquired alphabet. This suggestion is buttressed by the lack of a correlation between third-grade ambiguity scores in Roman and in Cyrillic. Apparently, for third graders the basis for the ambiguity effect in the two alphabets was not the same.

There are two possible reasons for the slower responses to ambiguous letter strings, in particular to the ambiguous Roman letter strings. One possible reason, anticipated when selecting children at two grade levels, is an overall Cyrillic bias when interpreting letter strings that is due to unequal proficiency with the two alphabets. This bias was restricted to ambiguous letter strings, however. Those letter strings that included unique letters were no slower in Roman than in Cyrillic at either level of alphabet proficiency. Moreover, the latency scores for the third graders who learned the Roman alphabet first revealed the same pattern. Nevertheless, a Cyrillic bias is suggested by the latency data on ambiguous words for third graders and is further supported by the alphabet-by-grade interaction in the analysis of variance on wrong alphabet errors. The other possible reason is an effect of two phonological analyses (where permitted) when proficiency with the two alphabets is equated. This possibility assumes equivalent performance with Roman and Cyrillic letter strings. The suggestion, therefore, is that a large ambiguity effect on Roman letter strings occurred for some third graders not because they were proficient decoders but, rather, because they were unfamiliar with the Roman alphabet. Put differently, the ambiguity effect with Roman materials that was manifested by third graders could have originated from one of two factors, where the two factors relate in opposite ways to reading skill. By contrast, fifth graders performed equivalently with ambiguous forms and unique alphabet controls in both the Roman and Cyrillic alphabets. As such, the fifth-grade data indicate a relation between reading skills and sensitivity to phonological ambiguity when the assumption of proficiency in the two alphabets is met.

Finally, one other difference between the skilled and less skilled beginning reader should be noted. The less skilled beginning reader of Serbo-Croatian is constrained by the fact that the characters of the Roman and Cyrillic alphabets belong to independent symbol systems. Table 7 reports the correlation of decoding speed with the tendency to mix alphabets in interpreting an ambiguous word. The less skilled decoder who has mastered both alphabets equally is more apt to ignore the independence of the two alphabets and to construct part of a word's name on the basis of a Roman alphabet reading and part on the basis of a Cyrillic alphabet reading.

In summary, the sequential acquisition of two alphabets in the process of learning to read, in conjunction with some special properties of the Serbo-Croatian language, permitted an investigation of the facility of beginning readers with a special variety of phonological analysis. It was demonstrated with children in the third and fifth grades that naming is slower and less accurate when a letter string can be assigned two phonological interpretations than when it can be assigned only one. This effect of ambiguity was assessed on two forms of the same word and it replicates earlier results with adults (Feldman, 1981; Feldman et al., 1983; Feldman & Turvey, 1983).

Two levels of bialphabetic profiency were examined. The asymmetry of the effect of phonological bivalence for ambiguous words in Roman as compared with Cyrillic was reduced as proficiency in each alphabet became equal and this suggested an analogy with the interaction of the two linguistic codes of the bilingual. For third graders, evidence of a Cyrillic bias when analyzing ambiguous letter strings made assessment of the relation between reading skill and sensitivity to phonological ambiguity equivocal. For fifth graders who are almost equally proficient in the two alphabets, however, there was evidence that the more skilled beginning reader was *more* impaired by phonological ambiguity than the less skilled beginning reader. In conclusion, the beginning reader who names letter strings more rapidly is more analytic in his or her style of reading than the poorer beginning reader.

References

Barron, R. W. (1978). Reading skill and phonological coding in lexical access. In M. M. Gruneberg, R. N. Sykes, & D. E. Morris (Eds.), *Practical aspects of memory*. London: Academic Press.

Coltheart, M. (1978). Lexical access in simple reading task. In G. Underwood (Ed.), *Strategies of information processing*. London: Academic Press.

Feldman, L. B. (1981). Visual word recognition in Serbo-Croatian is necessarily phonological. *Haskins Laboratories Status Report on Speech Research*,

Feldman, L. B. (1983). Bi-alphabetism and word recognition. *Published proceedings NATO Conference on the Acquisition of Symbolic Skills*. New York: Plenum Press.

Feldman, L. B., Kostić, A., Lukatela, G., & Turvey, M. T. (1983). An evaluation of the "Basic Orthographic Syllabic Structure" in a phonologically shallow orthography. *Psychological Research*, *45*, 55-72.

Feldman, L. B., & Turvey, M. T. (1983). Visual word recognition in Serbo-Croatian is phonologically analytic. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 288-298.

Hall, J. W., Wilson, K. P., Humphreys, M. S., Tinzmann, M. B., & Bowyer, P. M. (1983). Phonemic-similarity effects in good vs. poor readers. *Memory & Cognition*, *11*, 520-527.

Jackson, M. D. (1980). Further evidence for a relationship between memory access and verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, *19*, 683-694.

Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General*, *108*, 151-181.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. (1980). Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, *8*, 124-132.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. (1978). Bi-alphabetical lexical decision. *Language and Speech*, *21*, 142-165.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. (1978). On the relation between processing the Roman and Cyrillic alphabets: A preliminary analysis with bi-alphabetic readers. *Language and Speech*, *21*, 113-141.

Lukić, V. (1970). Active written vocabulary of pupils at the elementary school age (in Serbo-Croatian). Belgrade: Zavod za Izdavanje Udzbenika SR Srbije.

Magiste, E. (1984). Stroop tasks and dichotic translation: The development of interference patterns in bilinguals. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 304-325.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Martin, R. C. (1982). The pseudohomophone effect: The role of visual similarity in non-word decision. Quarterly Journal of Experimental Psychology, 34A, 395-409.

Perfetti, C. A., & Hogaboam, T. (1975). The relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 67, 461-469.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 5, 531-545.

Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), Orthographies and reading. London: Erlbaum.

Wolford, G., & Fowler, C. A. (1984). Differential use of partial information by good and poor readers. Developmental Review, 4.

Woodcock, R. W. (1973). Woodcock Reading Mastery Tests. Circle Pines, MN: American Guidance Services, Inc.

Footnotes

[1] This analytic style may be specific to language or it may be more general, embracing both linguistic and non-linguistic perception (see Wolford, & Fowler, 1984).

[2] In one failure to replicate these results (Hall, Wilson, Humphreys, Tinzmann, & Bowyer, 1983), a criterion for selecting good and poor readers was the math achievement test score from the Woodcock-Johnson battery (Woodcock, 1973). This test includes a subtest where word problems are presented orally so that successful performance on that test must involve short-term memory abilities. By constraining selection procedures in this way, all children with short-term memory problems were effectively eliminated from the Hall et al. study.

[3] For example, EKCEP can be interpreted either as /ekser/, which means "nail," or as /ektsep/, which is meaningless. The first form is based on a Cyrillic reading of EKCEP and the second on a Roman reading. By contrast, EKSER can only be interpreted in Roman, i.e., /ekser/. Therefore, EKCEP is phonologically ambiguous and EKSER is phonologically unique. The phonological representation associated with lexical access is sometimes termed morphophonological.

[4] For example, the possible unique alphabet transitions of BOPAM were БОПАМ (in Cyrillic) and VORAM (in Roman) where neither option was lexical. Therefore, alphabet designation for the unique alphabet transition of ambiguous pseudowords was randomly assigned and balanced over items.

VERTICALITY UNPARALLELED[*]

Ignatius G. Mattingly[†] and Alvin M. Liberman[††]

Having long found reason to believe that speech is special, we have, naturally enough, been surprised at the firmness with which others have asserted, to the contrary, that speech is just like everything else or, what comes to the same thing, that everything else is special, too. Apparently, our claim has run counter to some deeply-held conviction about the nature of mind. One of Fodor's achievements is that he makes this conviction explicit. On the orthodox view, as Fodor sees it, mental activities are "horizontally" organized; arguments for the specialness of speech and language fit better with the assumption that they are vertical. Of the many observations provoked by Fodor's lucid analysis of these opposing views, we can here offer only two. The first has to do with the relations among vertically organized input systems; the second, with the relations between input systems and output systems.

Fodor's input systems, being "domain specific" (p. 47), are in parallel, and their outputs complement each other. Thus, when two modules are sensitive to the same aspects of a signal, representations from both modules should be cognitively registered. This assumption is surely plausible for modules, such as those for shape and color, that compute complementary representations of the same distal object. But the situation is different for speech. There, the linguistic module appears to take precedence over the module (or modules) that look after distal objects that are not linguistic. Given the same aspect of the signal, the linguistic and the non-linguistic module are able to compute representations of different distal objects, but if a linguistic representation is computed, the non-linguistic representation is not cognitively registered. Consider an example to which Fodor himself alludes (p. 49): the transition of the third formant during the release of a consonantal constriction in a consonant-vowel syllable. When artificially isolated from the rest of the signal, this transition is perceived non-linguistically, as a chirp or glissando (Mann & Liberman, 1983; Repp, Milburn, & Ashkenas, 1983). But in its normal acoustic context, the same transition is not so heard; it simply contributes to the perception of a distal object that is distinctly linguistic: the place of articulation of the consonant.

---

[*] Invited commentary on J. A. Fodor, *The Modularity of Mind* (Cambridge, MA: MIT Press, 1983) to appear in *The Behavioral and Brain Sciences*.

[†] Also Department of Linguistics, University of Connecticut, Storrs, CT.

[††] Also Department of Psychology, University of Connecticut, Storrs, CT and Department of Linguistics, Yale University, New Haven, CT.

Acknowledgment. Support from NICHD Grant HD-01994 is gratefully acknowledged.

[HASKINS LABORATORIES: Status Report on Speech Research SR-79/80 (1984)]

Fodor's account of these facts would be that the isolated transition is ignored by the linguistic module, but not by the non-linguistic module, which registers it cognitively as a chirp. His account would also exclude the possibility that, for the transition in context, the linguistic module would register a chirp as well as a consonant. For the linguistic module, such a representation would be at most "intermediate" (p. 55 ff.), and hence inaccessible to central cognitive processes. (We ourselves doubt that the linguistic module computes any such representation at all, preferring to believe, instead, that the earliest representation is an articulatory one.) But the simple parallel arrangement of the modules that Fodor assumes does cause trouble, for while it means that "the computational systems that come into play in the perceptual analysis of speech...operate only upon acoustic signals that are taken to be utterances" (p. 49), it does not preclude the possibility that other systems will operate on these same signals. It suggests that the transition in context will be registered not only phonetically, by the linguistic module, but also non-phonetically, by the non-linguistic module. The listener would, therefore, hear both consonant and chirp. More generally, and more distressingly, the listener would hear all speech signals both as speech and as non-speech.

What seems called for is a mechanism that would guarantee the precedence of speech, but would not constitute a serious weakening of the modularity hypothesis. This precedence mechanism would insure that, though both the linguistic and the non-linguistic modules may be active (since speech and non-speech may occur simultaneously in the world), a signal will be heard as speech if possible, and otherwise as non-speech, but not as both. It is rather compelling evidence for the existence of such a mechanism that it can be defeated under experimental conditions that evade ecological constraints. This is what occurs in the phenomenon known as "duplex perception" (Liberman, Isenberg, & Rakerd, 1981; Mann & Liberman, 1983; Rand, 1974). As we have noted, if a third-formant transition that unambiguously fixes the perception of a consonant-vowel syllable (for example, either as /da/ or as /ga/) is extracted and presented in isolation, it sounds like a non-speech chirp. The remainder of the acoustic pattern, presented in isolation, is perceived as a consonant-vowel syllable, but in the absence of the transition, the place of the consonant is ambiguous. When the transition and the remainder are presented dichotically, a duplex percept results: the chirp is heard at the ear to which the transition is presented and an unambiguous consonant (/da/ or /ga/, depending on the transition) at the other ear; the ambiguous remainder is not also heard (Repp et al., 1983). Thus, the transition is perceived, simultaneously, as a non-speech chirp and as critical support for the consonant. Apparently, the precedence mechanism recognizes that the transition and the remainder belong together, but is also aware that there are two signal sources, one at each ear, and that only one of them is speech. It therefore allows both the linguistic module and the non-linguistic module to register central representations that depend on the formant transition.

How might this precedence mechanism work? An obvious possibility is that it scans the acoustic input and sorts speech signals from non-speech signals, routing each to its appropriate module. But such a sorting mechanism would seriously compromise the modularity view, because, having to cut across linguistic and non-linguistic domains, it would be blatantly horizontal. Fortunately, for the vertical view, the horizontal compromise appears to be wrong on empirical grounds.

The point is that a sorting mechanism would require that there be surface properties of speech that it could exploit. These properties would be characteristic of speech signals in general, but not of non-speech signals. Moreover, they would be distinct from those deeper properties that the linguistic module uses to determine phonetic structure. It is of considerable interest, then, that while natural speech signals do have certain surface properties (waveform periodicity, characteristic spectral structure, syllabic rhythm) that such a mechanism might be supposed to exploit (and that man-made devices for speech detection do exploit) none of these properties is essential for a signal to be perceived as speech. Natural speech remains speech-like, and even more or less intelligible, under many forms of distortion that destroy these properties (high- and low-pass filtering, infinite peak clipping, rate-adjustment). And, more tellingly, quite bizarre methods of synthesis--for example, replacing the formants of a natural utterances by sine waves with the same trajectories (Remez, Rubin, Pisoni, & Carrell, 1981)--suffice to produce speech-like signals. Thus, speech appears to be speech, not because of any surface properties that mark it as such, but entirely by virtue of properties that are deeply linguistic. A signal is speech if, and only if, the language module can in some degree interpret the signal as the result of phonetically significant vocal-tract gestures. (In the same way, there are no surface properties that distinguish grammatical sentences from ungrammatical ones: a sentence is grammatical if, and only if, a grammatical derivation can be given for it.) We therefore reject this horizontal compromise, and consider two other possible precedence mechanisms, both thoroughly vertical.

The first is an inhibitory precedence mechanism that works across the outputs of the modules in this way: If the linguistic module fails to find phonetic structure, then the output of the non-linguistic module is fully registered; if, on the other hand, the linguistic module does find phonetic structure, the link to the non-linguistic module causes the "corresponding" parts of its output to be inhibited, but leaves the phonetically irrelevant parts unaffected. Such a mechanism is certainly conceivable, and, being a central mechanism, would not compromise modularity. It would, however, be most unparsimonious. For if the inhibitory mechanism were to know which aspect of the output of the non-linguistic module corresponded to aspects of the signal that were treated as speech by the linguistic module, it would have to know everything that the two modules know: the relationships between phonetic structure and speech signals, as well as the relationship between non-linguistic objects and non-speech signals. Thus a central mechanism would, in effect, duplicate mechanisms of two of the modules.

Turning, therefore, to the second possible precedence mechanism, we propose that, while the outputs that modules provide to central processes are in parallel, their inputs may be in series. That is, one module may filter or otherwise transform the input signal to another module. We suppose that the linguistic module not only tracks the changing configuration of the vocal tract, recovering phonetic structure, but also filters out whatever in the signal is due to this configuration, including, of course, formant transitions. What remains--non-linguistic aspects of speech such as voice quality, loudness, and pitch, as well as unrelated acoustic signals--is passed on to the non-linguistic module. This supposition is parsimonious, in that it in no way complicates the computations we must attribute to the linguistic module; the information needed to perform the filtering is the same information that is needed to specify the phonetic structure of utterances (and ultimately the rest of their linguistic structure) to central processes.

A further point in favor of this serial precedence mechanism is that something similar appears to be required to explain the operation of other obvious candidates for module-hood, such as auditory localization, echo suppression and binocular vision. Consider just the first of these. The auditory localization module cannot simply be in parallel with other modules that operate on acoustic signals. Not only do we perceive sound sources (whether speech or non-speech) as localized (with the help of the auditory localization module), but we also fail to perceive unsynchronized left- and right-ear images (with other modules). Obviously, the auditory localization module does not merely provide information about sound-source locations to central cognitive processes; it also provides subsequent modules in the series, including the linguistic module, with a set of signals arrayed according to the location of their sources in the auditory field. The information needed to create this array (the difference in time-of-arrival of the various signals at the two ears) is identical to the information needed for localization.

Unfortunately, hypothesizing a serial precedence mechanism does not lead us directly to a full understanding of duplex perception. Until we have carried out some more experiments, we can only suggest that this phenomenon may have something to do with the fact that the linguistic module must not only separate speech from non-speech, but must also separate the speech of one speaker from that of another. For the latter purpose, it cannot rely merely on the differences in location of sound sources in the auditory field, since two speakers may occupy the same location, but must necessarily exploit the phonetic coherence within the signal from each speaker and the lack of such coherence between signals from different speakers. It might, in fact, analyze the phonetic information in its input array into one or more coherent patterns without relying on location at all, for under normal ecological conditions, there is no likelihood of coherence across locations. Thus, when a signal that is not in itself speech (the transition) nevertheless coheres phonetically with speech signals from a different location (the remainder of the consonant-vowel syllable), the module is somehow beguiled into using the same information twice, and duplex perception results.

Our second general observation about Fodor's essay is prompted by the fact that language is both an input system and an output system. Fodor devotes most of his attention to input systems and makes only passing mention (p. 42) of such output systems as those that may be supposed to regulate locomotion and manual gestures. He thus has no occasion to reflect on the fact that language is both perceptual and motoric. Of course, other modular systems are also in some sense both perceptual and motoric, and superficially comparable, therefore, to language: simple reflexes, for example, or the system that automatically adjusts the posture of a diving gannet in accordance with optical information specifying the distance from the surface of the water (Lee & Reddish, 1981). But such systems must obviously have separate components for detecting stimuli and initiating responses. It would make no great difference, indeed, if we chose to regard a reflex as an input system hard-wired to an output system, rather than as a single "input-output" system. What makes language (and perhaps some other animal communication systems also) of special interest is that, while the system has both input and output functions, we would not wish to suppose that there were two language modules, or even that there were separate input and output components within a single module. Assuming nature to have been a good communications engineer, we must rather suppose that there is but one module, within which corresponding input and output operations (parsing and sentence-planning; speech perception and speech production) rely on the same grammar, are computationally similar, and are executed by the same components. Computing logical form, given articulatory movements, and computing articulatory movements, given logical form, must somehow be the same process.

If this is the case, it places a strong constraint on our hypotheses about the nature of these internal operations. By no means every plausible account of language input is equally plausible, or even coherent, as an account of language output. The right kind of model would resemble an electrical circuit, for which the same system equation holds no matter where in the circuit we choose to measure "input" and "output" currents.

If the same module can serve both as part of an input system and part of an output system, the difference being merely a matter of transducers, then the distinction between perceptual faculties and motor faculties (the one fence Fodor hasn't knocked down) is perhaps no more fundamental than other "horizontal" distinctions. The fact that a particular module is perceptual, or motoric, or both, is purely "syncategoramatic" (p. 15). If so, then the mind is more vertical than even Fodor thinks it is.

References

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. Nature, 293, 293-294.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. Perception & Psychophysics, 30, 133-143.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. Cognition, 14, 211-235.

Rand, T. C. (1974). Dichotic release from masking for speech. Journal of the Acoustical Society of America, 55, 678-680.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. Science, 212, 947-950.

Repp, B. H., Milburn, C., & Ashkenas, J. (1983). Duplex perception: Confirmation of fusion. Perception & Psychophysics, 33, 333-337.

PUBLICATIONS
APPENDIX

PUBLICATIONS

Abramson, A. S. (1984). Obituary for Dennis Butler Fry. *Speech Communication*, *3*, 167-168.

Browman, C. P., & Goldstein, L. M. (in press). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), *Phonetic linguistics*. New York: Academic Press.

Feldman, L. B., Lukatela, G., & Turvey, M. T. (in press). Effects of phonological ambiguity in beginning readers of Serbo-Croatian. *Journal of Experimental Child Psychology*.

Fischer, F. W., Shankweiler, D., & Liberman, I. Y. (in press). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language*.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative, and Comparative*, *15*, R1000-R1004.

Kelso, J. A. S., Tuller, B., Bateson, E.-V., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 812-832.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E., & Kay, B. (in press). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America*.

Kidd, G. R., Boltz, M., & Jones, M. R. (1984). Some effects of rhythmic context on melody recognition. *American Journal of Psychology*, *97*,153-173.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. (1984). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren, Jr. & R. E. Shaw (Eds.), *Persistence and change: Proceedings of the First International Conference on Event Perception*. Hillsdale, NJ: Erlbaum.

Lisker, L., & Baer, T. (1984). Laryngeal management at utterance-internal word boundary in American English. *Language and Speech*, *27*, 163-171.

Löfqvist, A., McGarr, N. S., & Honda, K. (1984). Laryngeal muscles and articulatory control. *Journal of the Acoustical Society of America*, *76*, 951-954.

Mann, V. A. (1984). Reading skill and language skill. *Developmental Review*, *4*, 1-15.

Mann, V. A. (1984). Temporary memory for linguistic and nonlinguistic material in relation to the acquisition of Japanese Kana and Kanji. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, *18*, 127-134.

Mann, V. A. (1984). Perception of [l] and [r] by native speakers of Japanese: A distinction between articulatory tracking and phonetic categorization. *Annual Bulletin Research Institute of Logopedics and Phoniatrics*, *18*, 127-134.

Mattingly, I. G., & Liberman, A. M. (forthcoming). Verticality unparalleled (Review of *The modularity of the mind*, by J. A. Fodor. Cambridge, MA: MIT Press). *The Behavioral and Brain Sciences*.

Rakerd, B., Verbrugge, R. R., & Shankweiler, D. (1984). Monitoring for vowels in isolation and in a consonantal context. *Journal of the Acoustical Society of America*, *76*, 27-31.

Repp, B. H. (1984). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and Speech*, *27*, 245-254.

Repp, B. H., & Bentin, S. (in press). Parameters of spectral/temporal fusion in speech perception. *Perception & Psychophysics*.

Repp, B. H., & Williams, D. R. (in press). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. *Speech Communication*.

Saltzman, E., & Kelso, J. A. S. (in press). Synergies: stabilities, instabilities, and modes. A commentary on target article by L. M. Nashner and G. McCollum. *The Behavioral and Brain Sciences*.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*, *16*, 285-303.

Shankweiler, D., Smith, S. T., & Mann, V. A. (in press). Repetition and comprehension of spoken sentences by reading-disabled children. *Brain and Language*.

Tuller, B. (1984). On categorizing aphasic speech errors. *Neuropsychologia*, *22*, 547-557.

Tuller, B. (1984). Review of *Neuropsychology in reading, spelling, and language*, edited by U. Kirk. Invited review *Neuropsychologia*, *22*, 386-387.

Tuller, B., & Kelso, J. A. S. (1984). On the relative timing of articulatory gestures: Evidence for relational invariants. *Journal of the Acoustical Society of America*, *76*, 1030-1036.

APPENDIX

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834
SR-71/72	July - December 1982	AD A124596	ED 225-212
SR-73	January - March 1983	AD A129713	ED 229-816
SR-74/75	April - September 1983	AD A136416	ED 236-753
SR-76	October - December 1983	AD A140176	ED 241-973
SR-77/78	January - June 1984	AD A145585	ED 247-626
SR-79/80	July - December 1984	**	**

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*

Haskins Laboratories
270 Crown Street
New Haven, CT 06511

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

N/A

3. REPORT TITLE

Haskins Laboratories Status Report on Speech Research, SR-79/80 (1984)

4. DESCRIPTIVE NOTES *(Type of report and, inclusive dates)*

Interim Scientific Report

5. AUTHOR(S) *(First name, middle initial, last name)*

Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE

January, 1985

7a. TOTAL NO. OF PAGES

264

7b. NO. OF REFS

452

8a. CONTRACT OR GRANT NO.

HD-01994	NS13870
HD-16591	NS13617
N01-HD-1-2420	NS18010
RR-05596	N00014-83-K-0083
BNS-8111470	

9a. ORIGINATOR'S REPORT NUMBER(S)

SR-79/80

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES

N/A

12. SPONSORING MILITARY ACTIVITY

See No. 8

13. ABSTRACT

This report (1 July-31 December) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

- -Dynamic modeling of phonetic structure
- -Coarticulation as a component in articulatory description
- -Contextual effects on lingual-mandibular coordination
- -The timing of articulatory gestures: Evidence for relational invariants
- -Onset of voicing in stuttered and fluent utterances
- -Phonetic information is integrated across intervening nonlinguistic sounds
- -Parameters of spectral/temporal fusion in speech perception
- -Monitoring for vowels in isolation and in a consonantal context
- -Perception of [l] and [r] by native speakers of Japanese: A distinction between articulatory and phonetic perception
- -A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling
- -A theoretical note on speech timing
- -On reconciling monophthongal vowel percepts and continuously varying F patterns
- -Synergies: Stabilities, instabilities, and modes
- -Repetition and comprehension of spoken sentences by reading-disabled children
- -Spelling proficiency and sensitivity to word structure effects of phonological ambiguity on beginning readers of Serbo-Croatian
- -Effects of phonological ambiguity on beginning readers of Serbo-Croatian
- -Verticality unparalleled

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

Speech Perception:
  monophthongs, vowels, formants
  articulatory, phonetic, Japanese, [l]-[r]
  modular, perception, vertical
  consonantal context, monitoring
  spectral/temporal fusion
  integration, non-linguistic sound

Speech Articulation:
  voicing onset, stuttering, fluent
  models, dynamic, phonetic structure
  tongue, mandible, coordination
  coarticulation, kinematics, reiterant speech
  timing, phase, gestures, relational invariance

Motor Control:
  synergies, stabilities, instabilities, modes

Reading:
  beginning readers, phonology, ambiguity, Serbo-Croatian
  spelling, proficiency, sensitivity, word structure
  repetition, comprehension, reading-disabled, children
  spoken sentences